Software Testing
This module provides an introduction to software testing. Topics covered include basic definitions of testing, validation and verification; the levels of testing from unit testing through to acceptance testing; the relationship with requirements and design specifications; and test documentation.

## Introduction

Software Testing is the process of executing a program or system with the intent of finding errors. Or, it involves any activity aimed at evaluating an attribute or capability of a program or system and determining that it meets its required results. Software is not unlike other physical processes where inputs are received and outputs are produced. Where software differs is in the manner in which it fails. Most physical systems fail in a fixed (and reasonably small) set of ways. By contrast, software can fail in many bizarre ways. Detecting all of the different failure modes for software is generally infeasible.

Unlike most physical systems, most of the defects in software are design errors, not manufacturing defects. Software does not suffer from corrosion, wear-and-tear, generally it will not change until upgrades, or until obsolescence. So once the software is shipped, the design defects, or bugs, will be buried in and remain latent until activation.

Software bugs will almost always exist in any software module with moderate size: not because programmers are careless or irresponsible, but because the complexity of software is generally intractable, and humans have only limited ability to manage complexity. It is also true that for any complex systems, design defects can never be completely ruled out.

Discovering the design defects in software, is equally difficult, for the same reason of complexity. Because software and any digital systems are not continuous, testing boundary values are not sufficient to guarantee correctness. All the possible values need to be tested and verified, but complete testing is infeasible. Exhaustively testing a simple program to add only two integer inputs of 32-bits (yielding 2^64 distinct test cases) would take hundreds of years, even if tests were performed at a rate of thousands per second. Obviously, for a realistic software module, the complexity can be

far beyond the example mentioned here. If inputs from the real world are involved, the problem will get worse, because timing and unpredictable environmental effects and human interactions are all possible input parameters under consideration.

A further complication has to do with the dynamic nature of programs. If a failure occurs during preliminary testing and the code is changed, the software may now work for a test case that it didn't work for previously. But its behavior on pre-error test cases that it passed before can no longer be guaranteed. To account for this possibility, testing should be restarted. The expense of doing this is often prohibitive.

An interesting analogy parallels the difficulty in software testing with the pesticide (known as the Pesticide Paradox): Every method you use to prevent or find bugs leaves a residue of subtler bugs against which those methods are ineffectual. But this alone will not guarantee to make the software better, because the Complexity Barrier principle states: Software complexity (and therefore that of bugs) grows to the limits of our ability to manage that complexity. By eliminating the (previous) easy bugs you allowed another escalation of features and complexity, but his time you have subtler bugs to face, just to retain the reliability you had before. Society seems to be unwilling to limit complexity because we all want that extra bell, whistle, and feature interaction. Thus, our users always push us to the complexity barrier and how close we can approach that barrier is largely determined by the strength of the techniques we can wield against ever more complex and subtle bugs.

Regardless of the limitations, testing is an integral part in software development. It is broadly deployed in every phase in the software development cycle. Typically, more than 50% percent of the development time is spent in testing. Testing is usually performed for the following purposes:


**To improve quality**

As computers and software are used in critical applications, the outcome of a bug can be severe. Bugs can cause huge losses. Bugs in critical systems have

caused airplane crashes, allowed space shuttle missions to go awry, halted trading on the stock market, and worse. Bugs can kill. Bugs can cause disasters. In a computerized embedded world, the quality and reliability of software is a matter of life and death.

Quality means the conformance to the specified design requirement. Being correct, the minimum requirement of quality, means performing as required under specified circumstances. Debugging, a narrow view of software testing, is performed heavily to find out design defects by the programmer. The imperfection of human nature makes it almost impossible to make a moderately complex program correct the first time. Finding the problems and get them fixed, is the purpose of debugging in programming phase.

**For Verification & Validation (V&V)**

An important purpose of testing is verification and validation. Testing can serve as metrics. It is heavily used as a tool in the V&V process. Testers can make claims based on interpretations of the testing results, which either the product works under certain situations, or it does not work. We can also compare the quality among different products under the same specification, based on results from the same test.

We can not test quality directly, but we can test related factors to make quality visible. Quality has three sets of factors: functionality, engineering, and adaptability. These three sets of factors can be thought of as dimensions in the software quality space. Each dimension may be broken down into its component factors and considerations at successively lower levels of detail. Table 1 illustrates some of the most frequently cited quality considerations.

| Functionality(exterior quality) | Engineering(interior quality) | Adaptability(future quality) |
| --- | --- | --- |
|  |  |  |

| | | |
|---|---|---|
| Correctness | Efficiency | Flexibility |
| Reliability | Testability | Reusability |
| Usability | Documentation | Maintainability |
| Integrity | Structure | |

Table 1.  Typical Software Quality Factors

Good testing provides measures for all relevant factors. The importance of any particular factor varies from application to application. Any system where human lives are at stake must place extreme emphasis on reliability and integrity. In the typical business system usability and maintainability are the key factors, while for a one-time scientific program neither may be significant. Our testing, to be fully effective, must be geared to measuring each relevant factor and thus forcing quality to become tangible and visible.

Tests with the purpose of validating the product works are named clean tests, or positive tests. The drawbacks are that it can only validate that the software works for the specified test cases. A finite number of tests can not validate that the software works for all situations. On the contrary, only one failed test is sufficient enough to show that the software does not work. Dirty tests, or negative tests, refers to the tests aiming at breaking the software, or showing that it does not work. A piece of software must have sufficient exception handling capabilities to survive a significant level of dirty tests.

A testable design is a design that can be easily validated, falsified and maintained. Because testing is a rigorous effort and requires significant time and cost, design for testability is also an important design rule for software development.

**For reliability estimation**

Software reliability has important relations with many aspects of software, including the structure, and the amount of testing it has been subjected to.

Based on an operational profile (an estimate of the relative frequency of use of various inputs to the program), testing can serve as a statistical sampling method to gain failure data for reliability estimation.

Software testing is not mature. It still remains an art, because we still cannot make it a science. We are still using the same testing techniques invented 20-30 years ago, some of which are crafted methods or heuristics rather than good engineering methods. Software testing can be costly, but not testing software is even more expensive, especially in places that human lives are at stake. Solving the software-testing problem is no easier than solving the Turing halting problem. We can never be sure that a piece of software is correct. We can never be sure that the specifications are correct. No verification system can verify every correct program. We can never be certain that a verification system is correct either.

## Test Levels

### The target of the test

Software testing is usually performed at different levels along the development and maintenance processes. That is to say, the target of the test can vary: a single module, a group of such modules (related by purpose, use, behavior, or structure), or a whole system. Three big test stages can be conceptually distinguished, namely Unit, Integration, and System. No process model is implied, nor are any of those three stages assumed to have greater importance than the other two.

#### Unit testing

Unit testing verifies the functioning in isolation of software pieces which are separately testable. Depending on the context, these could be the individual subprograms or a larger component made of tightly related units. A test unit is defined more precisely in the IEEE Standard for Software Unit Testing (IEEE1008-87), which also describes an integrated approach to systematic and documented unit testing. Typically, unit testing occurs with access to the

code being tested and with the support of debugging tools, and might involve the programmers who wrote the code.

**Integration testing**

Integration testing is the process of verifying the interaction between software components. Classical integration testing strategies, such as top-down or bottom-up, are used with traditional, hierarchically structured software.

Modern systematic integration strategies are rather architecture-driven, which implies integrating the software components or subsystems based on identified functional threads. Integration testing is a continuous activity, at each stage of which software engineers must abstract away lower-level perspectives and concentrate on the perspectives of the level they are integrating. Except for small, simple software, systematic, incremental integration testing strategies are usually preferred to putting all the components together at once, which is pictorially called "big bang" testing.

**System testing**

System testing is concerned with the behavior of a whole system. The majority of functional failures should already have been identified during unit and integration testing. System testing is usually considered appropriate for comparing the system to the non-functional system requirements, such as security, speed, accuracy, and reliability. External interfaces to other applications, utilities, hardware devices, or the operating environment are also evaluated at this level.

## Objectives of Testing

Testing is conducted in view of a specific objective, which is stated more or less explicitly, and with varying degrees of precision. Stating the objective in

precise, quantitative terms allows control to be established over the test process.

Testing can be aimed at verifying different properties. Test cases can be designed to check that the functional specifications are correctly implemented, which is variously referred to in the literature as conformance testing, correctness testing, or functional testing. However, several other nonfunctional properties may be tested as well, including performance, reliability, and usability, among many others.

Other important objectives for testing include (but are not limited to) reliability measurement, usability evaluation, and acceptance, for which different approaches would be taken. Note that the test objective varies with the test target; in general, different purposes being addressed at a different level of testing.

References recommended above for this topic describe the set of potential test objectives. The sub-topics listed below are those most often cited in the literature. Note that some kinds of testing are more appropriate for custom-made software packages, installation testing, for example; and others for generic products, like beta testing.

**Qualification testing**

Qualification testing checks the system behavior against the customer's requirements, however these may have been expressed; the customers undertake, or specify, typical tasks to check that their requirements have been met or that the organization has identified these for the target market for the software. This testing activity may or may not involve the developers of the system.

**Installation testing**

Usually after completion of software and acceptance testing, the software can be verified upon installation in the target environment. Installation testing

can be viewed as system testing conducted once again according to hardware configuration requirements. Installation procedures may also be verified.

### Alpha and beta testing

Before the software is released, it is sometimes given to a small, representative set of potential users for trial use, either in-house (alpha testing) or external (beta testing). These users report problems with the product. Alpha and beta use is often uncontrolled, and is not always referred to in a test plan.

### Reliability achievement and evaluation

In helping to identify faults, testing is a means to improve reliability. By contrast, by randomly generating test cases according to the operational profile, statistical measures of reliability can be derived. Using reliability growth models, both objectives can be pursued together.

Software reliability refers to the probability of failure-free operation of a system. It is related to many aspects of software, including the testing process. Directly estimating software reliability by quantifying its related factors can be difficult. Testing is an effective sampling method to measure software reliability. Guided by the operational profile, software testing (usually black-box testing) can be used to obtain failure data, and an estimation model can be further used to analyze the data to estimate the present reliability and predict future reliability. Therefore, based on the estimation, the developers can decide whether to release the software, and the users can decide whether to adopt and use the software. Risk of using software can also be assessed based on reliability information. The primary goal of testing should be to measure the dependability of tested software.

There is agreement on the intuitive meaning of dependable software: it does not fail in unexpected or catastrophic ways. Robustness testing and stress testing are variances of reliability testing based on this simple criterion.

The robustness of a software component is the degree to which it can function correctly in the presence of exceptional inputs or stressful environmental conditions. Robustness testing differs with correctness testing in the sense that the functional correctness of the software is not of concern. It only watches for robustness problems such as machine crashes, process hangs or abnormal termination. The oracle is relatively simple, therefore robustness testing can be made more portable and scalable than correctness testing. This research has drawn more and more interests recently, most of which uses commercial operating systems as their target.

Stress testing, or load testing, is often used to test the whole system rather than the software alone. In such tests the software or system are exercised with or beyond the specified limits. Typical stress includes resource exhaustion, bursts of activities, and sustained high loads.

**Regression testing**

According to (IEEE610.12-90), regression testing is the "selective retesting of a system or component to verify that modifications have not caused unintended effects..." In practice, the idea is to show that software which previously passed the tests still does. Beizer defines it as any repetition of tests intended to show that the software's behavior is unchanged, except insofar as required. Obviously a trade-off must be made between the assurance given by regression testing every time a change is made and the resources required to do that.

Regression testing can be conducted at each of the test levels. The target of the test and may apply to functional and nonfunctional testing.

**Correctness testing**

Correctness is the minimum requirement of software, the essential purpose of testing. Correctness testing will need some type of oracle, to tell the right behavior from the wrong one. The tester may or may not know the inside details of the software module under test, e.g. control flow, data flow, etc.

Therefore, either a white-box point of view or black-box point of view can be taken in testing software. We must note that the black-box and white-box ideas are not limited in correctness testing only.

**Black-box testing**

The black-box approach is a testing method in which test data are derived from the specified functional requirements without regard to the final program structure. It is also termed data-driven, input/output driven, or requirements-based testing. Because only the functionality of the software module is of concern, black-box testing also mainly refers to functional testing - a testing method emphasized on executing the functions and examination of their input and output data. The tester treats the software under test as a black box - only the inputs, outputs and specification are visible, and the functionality is determined by observing the outputs to corresponding inputs. In testing, various inputs are exercised and the outputs are compared against specification to validate the correctness. All test cases are derived from the specification. No implementation details of the code are considered.

It is obvious that the more we have covered in the input space, the more problems we will find and therefore we will be more confident about the quality of the software. Ideally we would be tempted to exhaustively test the input space. But as stated above, exhaustively testing the combinations of valid inputs will be impossible for most of the programs, let alone considering invalid inputs, timing, sequence, and resource variables. Combinatorial explosion is the major roadblock in functional testing. To make things worse, we can never be sure whether the specification is either correct or complete. Due to limitations of the language used in the specifications (usually natural language), ambiguity is often inevitable. Even if we use some type of formal or restricted language, we may still fail to write down all the possible cases in the specification. Sometimes, the specification itself becomes an intractable problem: it is not possible to specify precisely every situation that can be encountered using limited words. And people can seldom specify clearly what they want - they usually can tell whether a prototype is, or is not, what they want after they have been

finished. Specification problems contributes approximately 30 percent of all bugs in software.

The research in black-box testing mainly focuses on how to maximize the effectiveness of testing with minimum cost, usually the number of test cases. It is not possible to exhaust the input space, but it is possible to exhaustively test a subset of the input space. Partitioning is one of the common techniques. If we have partitioned the input space and assume all the input values in a partition is equivalent, then we only need to test one representative value in each partition to sufficiently cover the whole input space. Domain testing partitions the input domain into regions, and consider the input values in each domain an equivalent class. Domains can be exhaustively tested and covered by selecting a representative value(s) in each domain. Boundary values are of special interest. Experience shows that test cases that explore boundary conditions have a higher payoff than test cases that do not. Boundary value analysis requires one or more boundary values selected as representative test cases. The difficulties with domain testing are that incorrect domain definitions in the specification can not be efficiently discovered.

Good partitioning requires knowledge of the software structure. A good testing plan will not only contain black-box testing, but also white-box approaches, and combinations of the two.

**White-box testing**

Contrary to black-box testing, software is viewed as a white-box, or glass-box in white-box testing, as the structure and flow of the software under test are visible to the tester. Testing plans are made according to the details of the software implementation, such as programming language, logic, and styles. Test cases are derived from the program structure. White-box testing is also called glass-box testing, logic-driven testing or design-based testing.

There are many techniques available in white-box testing, because the problem of intractability is eased by specific knowledge and attention on the structure of the software under test. The intention of exhausting some aspect of the software is still strong in white-box testing, and some degree of

exhaustion can be achieved, such as executing each line of code at least once (statement coverage), traverse every branch statements (branch coverage), or cover all the possible combinations of true and false condition predicates (Multiple condition coverage).

Control-flow testing, loop testing, and data-flow testing, all maps the corresponding flow structure of the software into a directed graph. Test cases are carefully selected based on the criterion that all the nodes or paths are covered or traversed at least once. By doing so we may discover unnecessary "dead" code - code that is of no use, or never get executed at all, which can not be discovered by functional testing.

In mutation testing, the original program code is perturbed and many mutated programs are created, each contains one fault. Each faulty version of the program is called a mutant. Test data are selected based on the effectiveness of failing the mutants. The more mutants a test case can kill, the better the test case is considered. The problem with mutation testing is that it is too computationally expensive to use. The boundary between black-box approach and white-box approach is not clear-cut. Many testing strategies mentioned above, may not be safely classified into black-box testing or white-box testing. It is also true for transaction-flow testing, syntax testing, finite-state testing, and many other testing strategies not discussed in this text. One reason is that all the above techniques will need some knowledge of the specification of the software under test. Another reason is that the idea of specification itself is broad - it may contain any requirement including the structure, programming language, and programming style as part of the specification content.

We may be reluctant to consider random testing as a testing technique. The test case selection is simple and straightforward: they are randomly chosen. Random testing is more cost effective for many programs. Some very subtle errors can be discovered with low cost. And it is also not inferior in coverage than other carefully designed testing techniques. One can also obtain reliability estimate using random testing results based on operational profiles. Effectively combining random testing with other testing techniques may yield more powerful and cost-effective testing strategies.

**Performance testing**

Not all software systems have specifications on performance explicitly. But every system will have implicit performance requirements. The software should not take infinite time or infinite resource to execute. "Performance bugs" sometimes are used to refer to those design problems in software that cause the system performance to degrade.

Performance has always been a great concern and a driving force of computer evolution. Performance evaluation of a software system usually includes: resource usage, throughput, stimulus-response time and queue lengths detailing the average or maximum number of tasks waiting to be serviced by selected resources. Typical resources that need to be considered include network bandwidth requirements, CPU cycles, disk space, disk access operations, and memory usage. The goal of performance testing can be performance bottleneck identification, performance comparison and evaluation, etc. The typical method of doing performance testing is using a benchmark - a program, workload or trace designed to be representative of the typical system usage.

**Security testing**

Software quality, reliability and security are tightly coupled. Flaws in software can be exploited by intruders to open security holes. With the development of the Internet, software security problems are becoming even more severe.

Many critical software applications and services have integrated security measures against malicious attacks. The purpose of security testing of these systems include identifying and removing software flaws that may potentially lead to security violations, and validating the effectiveness of security measures. Simulated security attacks can be performed to find vulnerabilities.

**Testing automation**

Software testing can be very costly. Automation is a good way to cut down time and cost. Software testing tools and techniques usually suffer from a lack of generic applicability and scalability. The reason is straight-forward. In order to automate the process, we have to have some ways to generate oracles from the specification, and generate test cases to test the target software against the oracles to decide their correctness. Today we still don't have a full-scale system that has achieved this goal. In general, significant amount of human intervention is still needed in testing. The degree of automation remains at the automated test script level.

The problem is lessened in reliability testing and performance testing. In robustness testing, the simple specification and oracle: doesn't crash, doesn't hang suffices. Similar simple metrics can also be used in stress testing.

**When to stop testing?**

Testing is potentially endless. We can not test till all the defects are unearthed and removed - it is simply impossible. At some point, we have to stop testing and ship the software. The question is when.

Realistically, testing is a trade-off between budget, time and quality. It is driven by profit models. The pessimistic, and unfortunately most often used approach is to stop testing whenever some, or any of the allocated resources - time, budget, or test cases - are exhausted. The optimistic stopping rule is to stop testing when either reliability meets the requirement, or the benefit from continuing testing cannot justify the testing cost. This will usually require the use of reliability models to evaluate and predict reliability of the software under test. Each evaluation requires repeated running of the following cycle: failure data gathering - modeling - prediction. This method does not fit well for ultra-dependable systems, however, because the real field failure data will take too long to accumulate.

**Alternatives to testing**

Software testing is more and more considered a problematic method toward better quality. Using testing to locate and correct software defects can be an endless process. Bugs cannot be completely ruled out. Just as the complexity barrier indicates: chances are testing and fixing problems may not necessarily improve the quality and reliability of the software. Sometimes fixing a problem may introduce much more severe problems into the system, happened after bug fixes.

Using formal methods to "prove" the correctness of software is also an attracting research direction. But this method can not surmount the complexity barrier either. For relatively simple software, this method works well. It does not scale well to those complex, full-fledged large software systems, which are more error-prone.

In a broader view, we may start to question the utmost purpose of testing. Why do we need more effective testing methods anyway, since finding defects and removing them does not necessarily lead to better quality. An analogy of the problem is like the car manufacturing process. In the craftsmanship epoch, we make cars and hack away the problems and defects. But such methods were washed away by the tide of pipelined manufacturing and good quality engineering process, which makes the car defect-free in the manufacturing phase. This indicates that engineering the design process (such as clean-room software engineering) to make the product have less defects may be more effective than engineering the testing process. Testing is used solely for quality monitoring and management, or, "design for testability". This is the leap for software from craftsmanship to engineering.

## Test Techniques

One of the aims of testing is to reveal as much potential for failure as possible, and many techniques have been developed to do this, which attempt to "break" the program, by running one or more tests drawn from identified classes of executions deemed equivalent. The leading principle underlying such techniques is to be as systematic as possible in identifying a representative set of program behaviors; for instance, considering subclasses of the input domain, scenarios, states, and dataflow.

It is difficult to find a homogeneous basis for classifying all techniques, and the one used here must be seen as a compromise. The classification is based on how tests are generated from the software engineer's intuition and experience, the specifications, the code structure, the (real or artificial) faults to be discovered, the field usage, or, finally, the nature of the application. Sometimes these techniques are classified as white-box, also called glassbox, if the tests rely on information about how the software has been designed or coded, or as black-box if the test cases rely only on the input/output behavior. One last category deals with combined use of two or more techniques. Obviously, these techniques are not used equally often by all practitioners. Included in the list are those that a software engineer should know.

## Based on the software engineer's intuition and experience

### Ad hoc testing

Perhaps the most widely practiced technique remains ad hoc testing: tests are derived relying on the software engineer's skill, intuition, and experience with similar programs. Ad hoc testing might be useful for identifying special tests, those not easily captured by formalized techniques.

### Exploratory testing

Exploratory testing is defined as simultaneous learning, test design, and test execution; that is, the tests are not defined in advance in an established test plan, but are dynamically designed, executed, and modified. The effectiveness of exploratory testing relies on the software engineer's knowledge, which can be derived from various sources: observed product behavior during testing, familiarity with the application, the platform, the failure process, the type of possible faults and failures, the risk associated with a particular product, and so on.

# Specification-based techniques

### Equivalence partitioning

The input domain is subdivided into a collection of subsets, or equivalent classes, which are deemed equivalent according to a specified relation, and a representative set of tests (sometimes only one) is taken from each class.

### Boundary-value analysis

Test cases are chosen on and near the boundaries of the input domain of variables, with the underlying rationale that many faults tend to concentrate near the extreme values of inputs. An extension of this technique is robustness testing, wherein test cases are also chosen outside the input domain of variables, to test program robustness to unexpected or erroneous inputs.

### Decision table

Decision tables represent logical relationships between conditions (roughly, inputs) and actions (roughly, outputs). Test cases are systematically derived by considering every possible combination of conditions and actions. A related technique is cause-effect graphing.

### Finite-state machine-based

By modeling a program as a finite state machine, tests can be selected in order to cover states and transitions on it.

### Testing from formal specifications

Giving the specifications in a formal language allows for automatic derivation of functional test cases, and, at the same time, provides a reference output, an oracle, for checking test results. Methods exist for deriving test cases from model-based or algebraic specifications.

**Random testing**

Tests are generated purely at random, not to be confused with statistical testing from the operational profile. This form of testing falls under the heading of the specification-based entry, since at least the input domain must be known, to be able to pick random points within it.

## Code-based techniques

**Control-flow-based criteria**

Control-flow-based coverage criteria is aimed at covering all the statements or blocks of statements in a program, or specified combinations of them. Several coverage criteria have been proposed, like condition/decision coverage. The strongest of the control-flow-based criteria is path testing, which aims to execute all entry-to-exit control flow paths in the flowgraph. Since path testing is generally not feasible because of loops, other less stringent criteria tend to be used in practice, such as statement testing, branch testing, and condition/decision testing. The adequacy of such tests is measured in percentages; for example, when all branches have been executed at least once by the tests, 100% branch coverage is said to have been achieved.

**Data flow-based criteria**

In data-flow-based testing, the control flowgraph is annotated with information about how the program variables are defined, used, and killed

(undefined). The strongest criterion, all definition-use paths, requires that, for each variable, every control flow path segment from a definition of that variable to a use of that definition is executed. In order to reduce the number of paths required, weaker strategies such as all-definitions and all-uses are employed.

**Reference models for code-based testing**

Although not a technique in itself, the control structure of a program is graphically represented using a flowgraph in code-based testing techniques. A flowgraph is a directed graph the nodes and arcs of which correspond to program elements. For instance, nodes may represent statements or uninterrupted sequences of statements, and arcs the transfer of control between nodes.

# Fault-based techniques

With different degrees of formalization, fault-based testing techniques devise test cases specifically aimed at revealing categories of likely or predefined faults.

**Error guessing**

In error guessing, test cases are specifically designed by software engineers trying to figure out the most plausible faults in a given program. A good source of information is the history of faults discovered in earlier projects, as well as the software engineer's expertise.

**Mutation testing**

A mutant is a slightly modified version of the program under test, differing from it by a small, syntactic change. Every test case exercises both the original and all generated mutants: if a test case is successful in identifying

the difference between the program and a mutant, the latter is said to be "killed." Originally conceived as a technique to evaluate a test set, mutation testing is also a testing criterion in itself: either tests are randomly generated until enough mutants have been killed, or tests are specifically designed to kill surviving mutants. In the latter case, mutation testing can also be categorized as a code-based technique. The underlying assumption of mutation testing, the coupling effect, is that by looking for simple syntactic faults, more complex but real faults will be found. For the technique to be effective, a large number of mutants must be automatically derived in a systematic way.

## Usage-based techniques

### Operational profile

In testing for reliability evaluation, the test environment must reproduce the operational environment of the software as closely as possible. The idea is to infer, from the observed test results, the future reliability of the software when in actual use. To do this, inputs are assigned a probability distribution, or profile, according to their occurrence in actual operation.

### Software Reliability Engineered Testing

Software Reliability Engineered Testing (SRET) is a testing method encompassing the whole development process, whereby testing is "designed and guided by reliability objectives and expected relative usage and criticality of different functions in the field."

## Techniques based on the nature of the application

The above techniques apply to all types of software. However, for some kinds of applications, some additional know-how is required for test

derivation. A list of a few specialized testing fields is provided here, based on the nature of the application under test:

- Object-oriented testing
- Component-based testing
- Web-based testing
- GUI testing
- Testing of concurrent programs
- Protocol conformance testing
- Testing of real-time systems
- Testing of safety-critical systems (IEEE1228-94)

## Selecting and combining techniques

### Functional and structural

Specification-based and code-based test techniques are often contrasted as functional vs. structural testing. These two approaches to test selection are not to be seen as alternative but rather as complementary; in fact, they use different sources of information and have proved to highlight different kinds of problems. They could be used in combination, depending on budgetary considerations.

### Deterministic vs. random

Test cases can be selected in a deterministic way, according to one of the various techniques listed, or randomly drawn from some distribution of inputs, such as is usually done in reliability testing. Several analytical and empirical comparisons have been conducted to analyze the conditions that make one approach more effective than the other.

# Test-related measures

Sometimes, test techniques are confused with test objectives. Test techniques are to be viewed as aids which help to ensure the achievement of test objectives. For instance, branch coverage is a popular test technique. Achieving a specified branch coverage measure should not be considered the objective of testing per se: it is a means to improve the chances of finding failures by systematically exercising every program branch out of a decision point. To avoid such misunderstandings, a clear distinction should be made between test-related measures, which provide an evaluation of the program under test based on the observed test outputs, and those which evaluate the thoroughness of the test set.

Measurement is usually considered instrumental to quality analysis. Measurement may also be used to optimize the planning and execution of the tests. Test management can use several process measures to monitor progress.

## Evaluation of the program under test (IEEE982.1-98)

### Program measurements to aid in planning and designing testing (IEE982.1-88)

Measures based on program size (for example, source lines of code or function points) or on program structure (like complexity) are used to guide testing. Structural measures can also include measurements among program modules in terms of the frequency with which modules call each other.

### Fault types, classification, and statistics (EEE1044-93)

The testing literature is rich in classifications and taxonomies of faults. To make testing more effective, it is important to know which types of faults could be found in the software under test, and the relative frequency with which these faults have occurred in the past. This information can be very useful in making quality predictions, as well as for process improvement.

**Fault density (IEEE982.1-88)**

A program under test can be assessed by counting and classifying the discovered faults by their types. For each fault class, fault density is measured as the ratio between the number of faults found and the size of the program

**Life test, reliability evaluation**

A statistical estimate of software reliability, which can be obtained by reliability achievement and evaluation, n be used to evaluate a product and decide whether or not testing can be stopped.

**Reliability growth models**

Reliability growth models provide a prediction of reliability based on the failures observed under reliability achievement and evaluation They assume, in general, that the faults that caused the observed failures have been fixed (although some models also accept imperfect fixes), and thus, on average, the product's reliability exhibits an increasing trend. There now exist dozens of published models. Many are laid down on some common assumptions, while others differ. Notably, these models are divided into failure-count and time-between-failure models.

## Evaluation of the tests performed

**Coverage/thoroughness measures (IEEE982.1-88)**

Several test adequacy criteria require that the test cases systematically exercise a set of elements identified in the program or in the specifications. To evaluate the thoroughness of the executed tests, testers can monitor the elements covered, so that they can dynamically measure the ratio between covered elements and their total number. For example, it is possible to

measure the percentage of covered branches in the program flowgraph, or that of the functional requirements exercised among those listed in the specifications document. Code-based adequacy criteria require appropriate instrumentation of the program under test.

**Fault seeding**

Some faults are artificially introduced into the program before test. When the tests are executed, some of these eeded faults will be revealed, and possibly some faults which were already there will be as well. In theory, depending on which of the artificial faults are discovered, and how many, testing effectiveness can be evaluated, and the remaining number of genuine faults can be estimated. In practice, statisticians question the distribution and representativeness of seeded faults relative to genuine faults and the small sample size on which any extrapolations are based. Some also argue that this technique should be used with great care, since inserting faults into software involves the obvious risk of leaving them there.

**Mutation score**

In mutation testing, the ratio of killed mutants to the total number of generated mutants can be a measure of the effectiveness of the executed test set.

**Comparison and relative effectiveness of different techniques**

Several studies have been conducted to compare the relative effectiveness of different test techniques. It is important to be precise as to the property against which the techniques are being assessed; what, for instance, is the exact meaning given to the term "effectiveness"? Possible interpretations are: the number of tests needed to find the first failure, the ratio of the number of faults found through testing to all the faults found during and after testing, or how much reliability was improved. Analytical and empirical comparisons

between different techniques have been conducted according to each of the notions of effectiveness specified above.

## Test Process

Testing concepts, strategies, techniques, and measures need to be integrated into a defined and controlled process which is run by people. The test process supports testing activities and provides guidance to testing teams, from test planning to test output evaluation, in such a way as to provide justified assurance that the test objectives will be met cost-effectively.

### Practical considerations

#### Attitudes/Egoless programming

A very important component of successful testing is a collaborative attitude towards testing and quality assurance activities. Managers have a key role in fostering a generally favorable reception towards failure discovery during development and maintenance; for instance, by preventing a mindset of code ownership among programmers, so that they will not feel responsible for failures revealed by their code.

#### Test guides

The testing phases could be guided by various aims, for example: in risk-based testing, which uses the product risks to prioritize and focus the test strategy; or in scenario-based testing, in which test cases are defined based on specified software scenarios.

#### Test process management (IEEE1074-97, IEEE12207.0-96:s5.3.9)

Test activities conducted at different levels must be organized, together with people, tools, policies, and measurements, into a well-defined process which is an integral part of the life cycle. In IEEE/EIA Standard 12207.0, testing is not described as a stand-alone process, but principles for testing activities are included along with both the five primary life cycle processes and the supporting process. In IEEE Std 1074, testing is grouped with other evaluation activities as integral to the entire life cycle.

**Test documentation and work products (IEEE829-98)**

Documentation is an integral part of the formalization of the test process. The IEEE Standard for Software Test Documentation (IEEE829-98) provides a good description of test documents and of their relationship with one another and with the testing process. Test documents may include, among others, Test Plan, Test Design Specification, Test Procedure Specification, Test Case Specification, Test Log, and Test Incident or Problem Report. The software under test is documented as the Test Item. Test documentation should be produced and continually updated, to the same level of quality as other types of documentation in software engineering.

**Internal vs. independent test team**

Formalization of the test process may involve formalizing the test team organization as well. The test team can be composed of internal members (that is, on the project team, involved or not in software construction), of external members, in the hope of bringing in an unbiased, independent perspective, or, finally, of both internal and external members. Considerations of costs, schedule, maturity levels of the involved organizations, and criticality of the application may determine the decision.

**Cost/effort estimation and other process measures (IEEE982.1-88)**

Several measures related to the resources spent on testing, as well as to the relative fault-finding effectiveness of the various test phases, are used by

managers to control and improve the test process. These test measures may cover such aspects as number of test cases specified, number of test cases executed, number of test cases passed, and number of test cases failed, among others.

Evaluation of test phase reports can be combined with root-cause analysis to evaluate test process effectiveness in finding faults as early as possible. Such an evaluation could be associated with the analysis of risks. Moreover, the resources that are worth spending on testing should be commensurate with the use/criticality of the application: different techniques have different costs and yield different levels of confidence in product reliability.

**Termination**

A decision must be made as to how much testing is enough and when a test stage can be terminated. Thoroughness measures, such as achieved code coverage or functional completeness, as well as estimates of fault density or of operational reliability, provide useful support, but are not sufficient in themselves. The decision also involves considerations about the costs and risks incurred by the potential for remaining failures, as opposed to the costs implied by continuing to test.

**Test reuse and test patterns**

To carry out testing or maintenance in an organized and cost-effective way, the means used to test each part of the software should be reused systematically. This repository of test materials must be under the control of software configuration management, so that changes to software requirements or design can be reflected in changes to the scope of the tests conducted.

The test solutions adopted for testing some application types under certain circumstances, with the motivations behind the decisions taken, form a test pattern which can itself be documented for later reuse in similar projects.

## Test Activities

Under this topic, a brief overview of test activities is given; as often implied by the following description, successful management of test activities strongly depends on the Software Configuration Management process.

### Planning

Like any other aspect of project management, testing activities must be planned. Key aspects of test planning include coordination of personnel, management of available test facilities and equipment (which may include magnetic media, test plans and procedures), and planning for possible undesirable outcomes. If more than one baseline of the software is being maintained, then a major planning consideration is the time and effort needed to ensure that the test environment is set to the proper configuration.

### Test-case generation

Generation of test cases is based on the level of testing to be performed and the particular testing techniques. Test cases should be under the control of software configuration management and include the expected results for each test.

### Test environment development

The environment used for testing should be compatible with the software engineering tools. It should facilitate development and control of test cases, as well as logging and recovery of expected results, scripts, and other testing materials.

### Execution

Execution of tests should embody a basic principle of scientific experimentation: everything done during testing should be performed and documented clearly enough that another person could replicate the results. Hence, testing should be performed in accordance with documented procedures using a clearly defined version of the software under test.

**Test results evaluation**

The results of testing must be evaluated to determine whether or not the test has been successful. In most cases, "successful" means that the software performed as expected and did not have any major unexpected outcomes. Not all unexpected outcomes are necessarily faults, however, but could be judged to be simply noise. Before a failure can be removed, an analysis and debugging effort is needed to isolate, identify, and describe it. When test results are particularly important, a formal review board may be convened to evaluate them.

**Problem reporting/Test log**

Testing activities can be entered into a test log to identify when a test was conducted, who performed the test, what software configuration was the basis for testing, and other relevant identification information. Unexpected or incorrect test results can be recorded in a problem-reporting system, the data of which form the basis for later debugging and for fixing the problems that were observed as failures during testing. Also, anomalies not classified as faults could be documented in case they later turn out to be more serious than first thought. Test reports are also an input to the change management request process.

**Defect tracking**

Failures observed during testing are most often due to faults or defects in the software. Such defects can be analyzed to determine when they were introduced into the software, what kind of error caused them to be created

(poorly defined requirements, incorrect variable declaration, memory leak, programming syntax error, for example), and when they could have been first observed in the software. Defect-tracking information is used to determine what aspects of software engineering need improvement and how effective previous analyses and testing have been.

References:

http://en.wikipedia.org/wiki/Software_testing,
http://ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-171Fall2003/CourseHome/,
http://www.cs.cornell.edu/courses/cs501/2008sp/,
http://www.comp.lancs.ac.uk/computing/resources/IanS/SE7/,
http://www.ee.unb.ca/kengleha/courses/CMPE3213/IntroToSoftwareEng.htm
, http://www.cs.kuleuven.ac.be/~dirk/ada-belgium/aia/contents.html#5,
http://www.softwareqatest.com/qatfaq1.html, etc...

## Hypotheses Testing - Examples.

**Example:**
We have tossed a coin 50 times and we got **k = 19 heads**. Should we accept/reject the hypothesis that **p** = 0.5, provided taht the coin is fair?
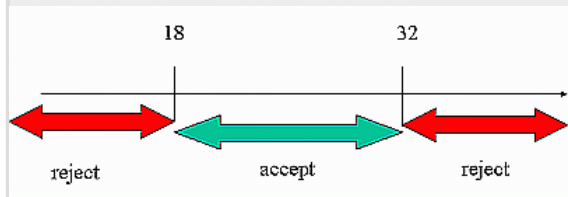**Null versus Alternative Hypothesis:**

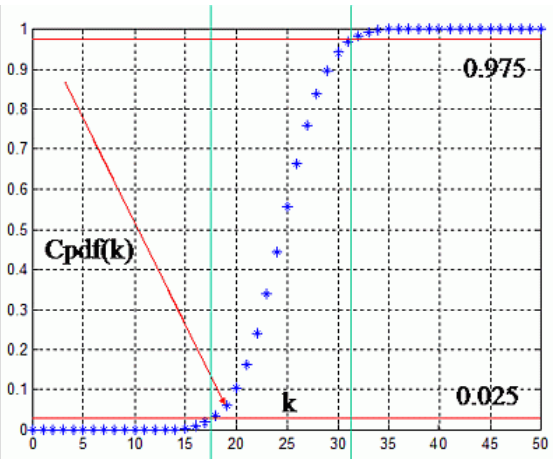- Null hypothesis
- Alternative hypothesis

**EXPERIMENT**
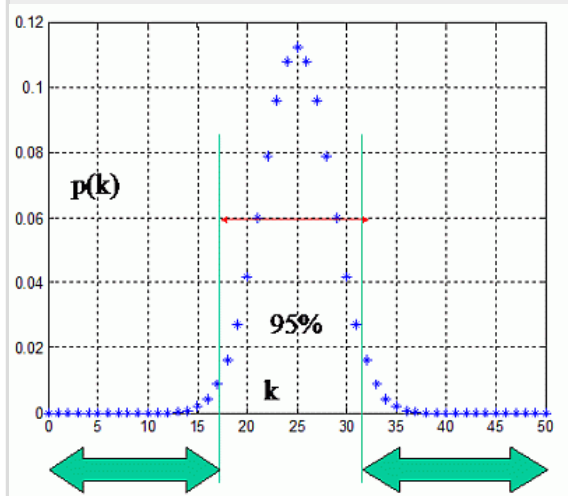


Significance level   = Probability of Type I error = Pr[rejecting      |      true]
**P[        or        ]        .**
If         or        ]        , then under the null hypothesis the observed event falls into rejection region with the probability

**Note:** We want    as small as possible.



Test construction.

Cumulative distribution function.

**Note:** No evidence to reject the null hypothesis.

**Example:**
We have tossed a coin 50 times and we got **k = 10 heads**. Should we accept/reject the hypothesis that **p** = 0.5, provided taht the coin is fair?
**EXPERIMENT**
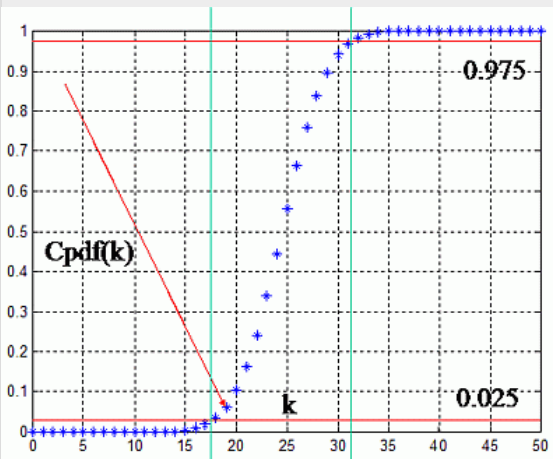


Cumulative distribution function.

**P[** **or** **]** We could **reject** hypothesis at a significance level as low as

**Note:** In STATISTICS, to prove something = **reject** the hypothesis that converse is true.

**Example:**
We know that on average mouse tail is 5 cm long. We have a group of 10 mice, and give to each of them a dose of vitamin **T** everyday, from the birth, for the period of 6 months.
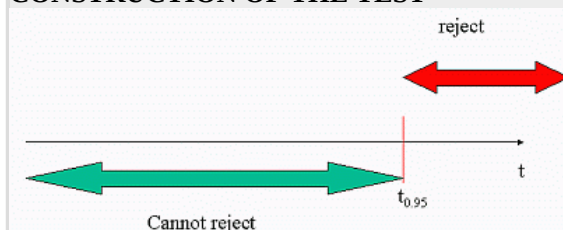We want to prove that vitamin **X** makes mouse tail longer. We measure tail lengths of out group and we get the following sample:

| 5.5 | 5.6 | 4.3 | 5.1 | 5.2 | 6.1 | 5.0 | 5.2 | 5.8 | 4.1 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Table 1

- Hypothesis    - sample = sample from normal distribution with    = 5 cm.
- Alternative    - sample = sample from normal distribution with    > 5 cm.

**CONSTRUCTION OF THE TEST**



We do not know population variance, and/or we suspect that vitamin treatment may change the variance - so we use t distribution.

- ⎯

    ⎯⎯⎯⎯⎯⎯⎯

- ⎯

- ⎯⎯    ⎯⎯⎯

**Example:**
   test (K. Pearson, 1900)

To test the hypothesis that a given data actually come from a population with the proposed distribution. Data is given in the Table 2.

| 0.4319 | 0.6874 | 0.5301 | 0.8774 | 0.6698 | 1.1900 | 0.4360 | 0.2192 | 0.5082 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.3564 | 1.2521 | 0.7744 | 0.1954 | 0.3075 | 0.6193 | 0.4527 | 0.1843 | 2.2617 |
| 0.4048 | 2.3923 | 0.7029 | 0.9500 | 0.1074 | 3.3593 | 0.2112 | 0.0237 | 0.0080 |
| 0.1897 | 0.6592 | 0.5572 | 1.2336 | 0.3527 | 0.9115 | 0.0326 | 0.2555 | 0.7095 |
| 0.2360 | 1.0536 | 0.6569 | 0.0552 | 0.3046 | 1.2388 | 0.1402 | 0.3712 | 1.6093 |
| 1.2595 | 0.3991 | 0.3698 | 0.7944 | 0.4425 | 0.6363 | 2.5008 | 2.8841 | 0.9300 |
| 3.4827 | 0.7658 | 0.3049 | 1.9015 | 2.6742 | 0.3923 | 0.3974 | 3.3202 | 3.2906 |
| 1.3283 | 0.4263 | 2.2836 | 0.8007 | 0.3678 | 0.2654 | 0.2938 | 1.9808 | 0.6311 |
| 0.6535 | 0.8325 | 1.4987 | 0.3137 | 0.2862 | 0.2545 | 0.5899 | 0.4713 | 1.6893 |
| 0.6375 | 0.2674 | 0.0907 | 1.0383 | 1.0939 | 0.1155 | 1.1676 | 0.1737 | 0.0769 |
| 1.1692 | 1.1440 | 2.4005 | 2.0369 | 0.3560 | 1.3249 | 0.1358 | 1.3994 | 1.4138 |
| 0.0046 | - | - | - | - | - | - | - | - |

DATA

**Exercise:**

**Problem:** Are these data sampled from population with exponential p.d.f.?

**Solution:**

**CONSTRUCTION OF THE TEST**
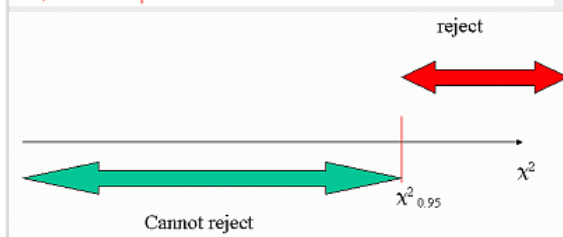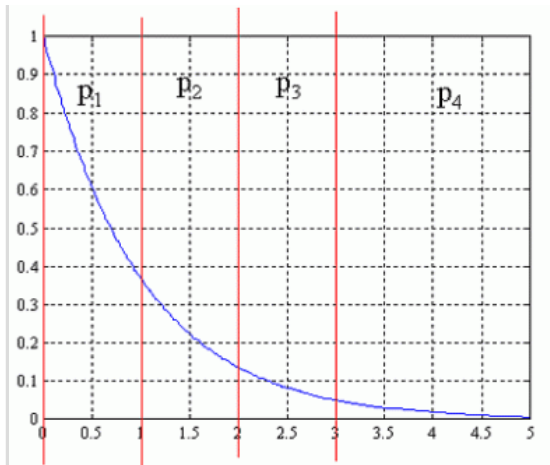
**Exercise:**

**Problem:** Are these data sampled from population with exponential p.d.f.?

**Solution:**

1. Estimate **a**.
2. Use     test.
3. Remember d.f. = K-2.

| Actual Situation | | true | | false | |
|---|---|---|---|---|---|
| decision | accept | Reject = error t. I | reject | | Accept = error t. II |
| probability | |  = significance level |  = power of the test | | |

TABLE 1

Oral Language Testing at Tay Nguyen University
Assessment of oral language proficiency at Tay Nguyen University (TNU), where the author of this thesis works, has been claimed to be extremely problematic. This thesis takes a critical look at the reality of oral English language testing at this institution to point out its strengths and weaknesses and the cause(s) of the existing drawbacks or problems.

## STATEMENT OF AUTHORSHIP

I certify that the minor thesis entitled 'Oral Language Testing at Tay Nguyen University: Current Practices and Recommendations for Improvement' and submitted in partial fulfilment of the requirements for the degree of Master of Arts in TESOL is the result of my work, except where otherwise acknowledged, and that this minor thesis or any part of the same has not been submitted for a higher degree to any other university or institution.

The research reported in this thesis was approved by Hanoi University of Foreign Studies.

Signed: Le Thi Phuong Nhi. DH Tay Nguyen

Dated:24 February 2008

## table of contents

## ACKNOWLEDGEMENTS

## glossary

This glossary is intended to give working definitions of terms used frequently in this thesis in order to help readers understand the author's intended meaning.

**Communicative stress** This term means the difficulty degree of a task that a speaker has to carry out. This difficulty refers to all the conditions under which the speaker is put to perform the task.

Elicitation technique An elicitation technique involves the procedure of performing a task that inferences of a speaker's language ability are based on.

Interactional function This term refers to one of the two functions of spoken language. A speaker producing an interactional instance of spoken language wants to make the interaction atmosphere pleasant.

Level scale or rating scaleA level or rating scale used in this thesis is a document displaying classified levels of learners' language knowledge and what learners can do at each level.

**Long turnA long turn is a string of utterances that a speaker produces**

Oral test type A type of oral test refers to the way a test task requires test takers to do.

Short turnA short turn is speech of one or two utterances that a speaker produces.

Test administration Test administration involves the delivery of a set of test tasks to a group of test takers under specified conditions.

Test designTest design in this thesis refers to the production of a principled statement as a basis for writing actual tests and administering them.

Test developmentTest development is the entire process of creating and using a test. It involves test design, test operationalization and test administration.

Test operationalizationTest operationalization is the production process of actual tests. It involves developing test task specifications and test structure.

Test structure Test structure refers to the number of test tasks included in a test

Test task specificationsTest task specifications tell in detail what a test is designed to measure and how it will be tested

Transactional functionThis term refers to one of the two functions of spoken language. A speaker producing a transactional instance of spoken language means to convey his intentions and messages.

## ABSTRACT

Assessment of oral language proficiency at Tay Nguyen University (TNU), where the author of this thesis works, has been claimed to be extremely problematic. This thesis takes a critical look at the reality of oral English language testing at this institution to point out its strengths and weaknesses and the cause(s) of the existing drawbacks or problems.

In order to achieve this, a study was carried out to evaluate TNU current practices and TNU staff's perceptions of oral language testing. The methods employed in the study include: (1) an detailed analysis of the current oral test development process based on Bachman & Palmer's theoretical framework for test development, and (2) a survey on how well the staff know about oral skill assessment.

The results of the study show that (1) the current oral testing practices at this institution are far from being consistent with the language testing theory, and (2) the staff have gained limited and insufficient knowledge of oral language testing. These findings serve as the basis for seven practical recommendations made for the improvement and standardisation of TNU current oral testing practices.

The seven recommendations are as follows. Recommendations 1,2,3,4 & 5 are made as an effort to make relevant applications which are based on Bachman & Palmer's theoretical framework for test development. These recommendations can be considered as guidelines for developing speaking tests in general. Recommendations 6 & 7 are particularly intended for the sample operationalization of speaking tests for first-year students.

**list of figures and tables**

**CHAPTER 1: INTRODUCTION**

This thesis reports the results of a study carried out to investigate the current practices of oral testing at Tay Nguyen University (TNU) in order to point out the existing problems and to make some practical suggestions for improvement. This introductory chapter describes in detail the problem the thesis attempts to solve, states the objectives of the study, and provides an overview of the thesis.

## 1.1 The Problem

### 1.1.1 Theoretical Perspective

Theoretically, as has become clear through empirical studies in language testing, there has been 'a shift from using assessment as a way to keep students in their place to using assessment as a way to help students find their place in school and in the world community of language users' (Cohen, 1996, p. 3). In this popular tendency of treating language tests, language tests have been considered extremely helpful for both students and teachers, and even for administrators. Madsen (1983, p. 4-5) points out the importance of language testing by demonstrating that properly made tests can

'1. help create positive attitudes towards instruction by giving students a sense of accomplishment and a feeling that the teacher's evaluation of them matches what he has taught them.

2. help students learn the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve.

3. help teachers and administrators by confirming progress that has been made and showing how they can best redirect their future efforts.'

Therefore, being competent in language testing, particularly in oral language testing under review in this thesis, is claimed to be crucial for language teachers to properly develop language tests. This thesis is an attempt to provide a clear discussion of how to become competent in oral language testing. An answer to this question will explicitly help to evaluate TNU current oral testing practices.

### 1.1.2 Practical Perspective

Apart from the above theoretical concern, this thesis also grows out of a practical consideration regarding the researcher's work at TNU as an English teacher and assessor of students' oral test performance. The problem identified in this thesis has taken root from the existing oral testing practices at TNU. The following are two tables of oral test results of the second-year students (School Year 2001-2002) and of the third-year students (School Year 2002-2003).

| Students with | Term 1 | Term 2 | Average |
|---|---|---|---|
| Mark 4 | 0% | 6,5% | 3,25% |
| Marks 5 and 6 | 38% | 41,5% | 39,75% |
| Mark 7 | 33,5% | 27% | 20,25% |
| Marks 8 and 9 | 28,5% | 25% | 26,75% |

Table 1.1: The second-year students' oral test results

| Students with | Term 1 | Term 2 | Average |
|---|---|---|---|
| Mark 4 | 0% | 0% | 0% |
| Marks 5 and 6 | 26% | 39% | 32,5% |
| Mark 7 | 40% | 33% | 36,5% |
| Marks 8 and 9 | 34% | 28% | 31% |

Table 1.2: The third-year students' oral test results

The two tables reveal that nearly half of the second-year students (47%) and more than half of the third-year students (67,5%) get high marks (7,8 and 9). However, in a talk with the researcher about their speaking ability, preferably their results of the former speaking tests, the majority of those students who got high marks in the tests seemed very reluctant to agree that their test results really revealed their actual ability to use English for communication. For example, they still found it hard to use English to either satisfactorily communicate their ideas or make themselves fully understood in a real instance of communication. So why did they get such high scores?

The same question was put forwards for discussion with the teachers and the assessors of speaking skill. Most of them believed that those students deserved high scores because they actually performed their tasks fluently and were able to answer the questions of the examiners. Nevertheless, when the author asked these teachers what particular speaking abilities they expected to see in the students' test performance, what exact criteria their assessment was based on, and what the detailed procedure for test design was, they did not all provided any specific and clear answers. They could not point out speaking abilities expected, they have assessed students' test performance intuitively, and they have constructed oral tests in their own way. Thus, as can be concluded that there seems to have never been any official detailed guidance for the construction and administration of oral tests at this institution.

Most of the staff members whom I have discussed this matter with have shared the same worry and showed interest in how to gain a scientific approach to assessing our students' oral ability properly and fairly. In other words, we all would like to find out how to appropriately measure the students' speaking ability and how to write useful speaking tests. This aims at helping to ensure fairness for the students, and improve and maintain the training quality of the institution.

All the concerns described above indicate an urgent need to evaluate the reality of oral language testing at TNU in the light of language testing theory. A thorough study is carried out as an attempt to help the staff give the oral testing an adequate position in their training program.

**1.2 Aims and Overview of the Thesis**

This thesis is carried out with the two main aims: firstly, to investigate the existing oral testing practices at TNU; and secondly, to make suggestions for improvement. Based on Bachman and Palmer's theoretical frammework for developing language tests, the author recommends a procedure for speaking test development as an attempt to provide a profound understanding of how to properly develop an oral test. Therefore, these recommendations are hopefully used as guidelines for oral language test development not only at TNU but also at other institutions throughout Vietnam. The data used for the two purposes above are collected from two different sources: (1) a detailed analysis of current oral testing practices at TNU, in particular, the development procedure of an

achievement speaking test; and (2) a questionnaire survey given out to 12 TNU teachers of English to investigate their perceptions of oral testing and thus to find out the cause(s) of the current practices.

The thesis consists of five chapters as follows:

Chapter 1 identifies the problem and provides an overview of the thesis.

Chapter 2 reviews the literature related to major issues in oral language testing such as essential features of spoken language, a theoretical framework for test development in general and development of oral tests in particular, and major qualities of a test.

Chapter 3 describes the methodology employed in the study. In order to evaluate current practices, the study involves describing the existing practices of spoken language testing at TNU, and investigating the staff's perceptions of oral testing by delivering questionnaires to 12 staff members.

Chapter 4 presents the results of the study, and analyses the results to point out findings.

Chapter 5 makes some practical recommendations for standardisation of TNU oral testing practices, and provides a summary of the main details of the whole thesis with a conclusion ending the thesis.

## CHAPTER 2: LITERATURE REVIEW

Chapter 1 presents the background information of the study. This chapter looks at main issues of oral testing. The discussion of the issues is meant to give a theoretical foundation on which to develop a framework for developing oral tests. The chapter discusses the following issues: (1) typical features of spoken language, (2) communicative approach to testing oral language ability, (3) theoretical framework for test development, (4) major considerations in construction of oral test tasks and tests, and (5) qualities of a good test.

### 2.1 Typical Features of Spoken Language

Spoken language had been ignored in language teaching long before it was noticed to be as essential as written language as well as other aspects of this science. From this time learners of a foreign language have been encouraged to learn how to produce spoken language forms spontaneously, not simply to utter written language sentences.

The features of spoken language reviewed here will help to specify typical and important areas of language knowledge to be involved in the process of testing speaking skill.

The most special feature of spoken language is its functions. Brown and Yule (1983) demonstrate that spoken language encompasses two functions in terms of a speaker's intention. These two functions are defined as Interactional function and Transactional function. The former refers to the kind of spoken language speakers use to make their interaction atmosphere pleasant whereas the latter is concerned in interactions where speakers want to mainly convey their intentions and messages. Therefore, Brown and Yule (1983, p.13) assert that interactional language is listener-oriented while transactional language is message-oriented.

In interactional situations the participating speakers do not challenge each other to communicate information, and tend to end up feeling friendly and comfortable with each other. In transactional situations information transmission requires language exchanges between interlocutors to be understandable and appropriate. Obviously, 'all foreign learners of English, who wish to learn the spoken form of the language, need to be able to express their transactional intentions'. They must know how to make clear the ideas to be communicated, even in their own mother tongue environment, yet it is easier to make themselves understood in their own language than in a new language.

Another crucial feature of spoken language is length of its production, that is the language is orally produced at length or not. Speech consisting of only one or two utterances is defined as a short turn, and that of a string of

utterances is defined as a long turn by Underhill (1987, p.16). Taking short turns is of course less demanding than taking long turns. When in position of taking a transactional long turn, a speaker is immediately 'responsible for creating a structured sequence of utterances which must help the listener(s) to create a coherent mental representation of what he is trying to say'.

As regards these two features a product of spoken language can be considered in such a continuum as interactional short turns – transactional short turns – transactional long turns. The difficulty of spoken language production ranges from the one extreme to the other extreme of the continuum, and the level of difficulty is shown in the figure 2.1 below. Clearly teaching as well as testing speaking skill should gradually follow this continuum according to learners' level of language proficiency.

Interactional Transactional

Short turns

Long turns

Figure 2.1: Continuum of Spoken Language Production

The above figure indicates that content to be taught or assessed should be graded according to the difficulty of tasks intended for the course purposes. The degree of this difficulty is determined by communicative stress, which involves three conditions under which a speaker feel more or less comfortable in producing what he has to (Brown and Yule, 1983, p.34). The less stressful a task is, the easier it is for speakers to carry out. These three conditions are features of the context, state of knowledge of the listener and type of task shown in the figure 2.2 below.

Communicative StressState of knowledge of the listenerType of taskFeatures of the context-The listener-The situation-The language-The information-Status of knowledge-Structure of the task

Figure 2.2: Conditions of Communicative Stress in a Task

The listener refers to the relationship between the speaker and the listener, or the number of the listeners he is talking to. The situation is concerned with the speaking environment (is it familiar or unfamiliar, and private or in public?). The language relates to the listener(s)' language proficiency in comparison with the speaker's, and the information is what the listener wants or needs. Status of knowledge mentions the degree of familiarity of the task's topic, and structure of the task refers to the purpose of the task or the difficulty of the task itself. This difficulty ranges from the static relationships to the abstract relationships between what is being talked about and what is going to be said. Obviously, tasks involving 'abstract relationships are more difficult than those involving the description of static and dynamic relationships' (Nunan, 1991, p. 48). O'Malley & Pierce (1996, p. 76) state these relationships correspond to an increase in difficulty levels. The tasks intended for the purpose(s) of teaching or testing should thus be graded according to these relationships as follows:

1. Static relationships

Describing an object or photograph

Instructing someone to draw a diagram

Instructing someone how to assemble a piece of equipment

Describing/instructing how a number of objects are to be arranged

Giving route directions

1. Dynamic relationships

Story-telling

Giving an eye-witness account

1. Abstract relationships

Opinion-expressing

Justifying a course of actions

(Brown & Yule, 1983, p.109)

The difficulty of tasks additionally depends upon the number of relationships, elements, factors or characters within each task. For instance, 'a short narrative involving a single character and only two or three events may be easier than a lengthy description covering many details and relationships'.

Linguistic KnowledgeSociocultural KnowledgeCooperative PrincipleProduction/Interpretation of Spoken LanguageFurthermore, how to ensure their production of spoken language in a new language to be appropriately interpreted is extremely demanding on the part of speakers/learners. In order to achieve this confidence, learners must first process their acquired knowledge of language and then produce utterances linguistically acceptable and socioculturally appropriate, and the utterances must conform to the cooperative principle (Celce-Murcia and Olshtain, 2000, p.168-171). This principle refers to the general rules of how to maintain the exchange flow between interlocutors, which means that 'the speaker wants to be understood and interpreted correctly and the hearer wants to be an effective decoder of the messages he receives'. A speaker's ideas successfully communicated are illustrated in figure 2.3 below.

Figure 2.3: Success of Meaning Negotiation

Linguistic knowledge, or Organizational knowledge (Bachman and Palmer, 1996; and Bachman, 1990), includes grammatical knowledge (ie. knowledge of vocabulary, morphology, syntax, and phonology/graphology), and textual knowledge (ie. rules of cohesion and coherence, and knowledge of rhetorical organisations). Sociocultural knowledge, or Pragmatic knowledge (Bachman and Palmer, 1996; and Bachman, 1990), is associated with '(1) characteristics of the individuals who take part in the communicative exchange, (2) features of the situation in which this exchange takes place, (3) the goal of the exchange, and (4) features of the communicative medium through which the exchange is carried out.

The assessment of learners'/students' production of spoken language or their oral test performance is entirely based on the two major features of spoken language - interactional and transactional functions, and production length. Therefore, the criteria for assessment must be formed and founded on the basis of these two features, and these criteria vary according to learners' language proficiency level or the difficulty of test tasks. In particular, the criteria for assessing learners' interactional and transactional short turns are to focus more on learners' or test takers' communicative reaction and successfully negotiated ideas rather than on content, size, cohesion or coherence like in taking transactional long turns.

To sum up, the two functions and length of spoken language production are deeply associated with what to be tested in a test of oral ability, and how to ensure success of spoken language production is primarily related to how to test or assess learners' oral ability. The next section will discusses the suitable approach to making right inferences from learners' oral test performance.

## 2.2 Communicative Approach to Testing Oral Language Ability

Testing the oral ability in a language is one of the most important aspects of language testing. This ability is an extremely difficult skill to assess as Heaton (1988) and Brown & Yule (1983) suppose. Partly because of the difficulty of treating speaking tests in the same way as other more conventional tests, testing of speaking skill has generally received little attention. In a genuine speaking test, real people meet face to face, and talk to each other. Hence, it is the people and what passes between them that are important whereas the test instrument is secondary. To put it more closely, oral tests should be designed around the people involved so that they can be encouraged to talk to each other as naturally as possible.

For several decades, a new theory of language and language use has exerted a considerable influence on language teaching and potentially on language testing. For example, Hymes's theory of communicative competence is concerned with not only language forms but also the ability to use language in socio-cultural context. Communicative competence in oral language 'requires control of a wide range of phonological and syntactic features, vocabulary, and oral genres and the knowledge of how to use them appropriately' (Butler et al., 2000, p.2) Although the relevance of this theory to language testing was recognized more or less immediately, it took quite long for its actual impact on practice to be felt in the development of communicative language tests. McNamara (2000, p. 16-17) characterises communicative language tests to have two features:

1. they are performance tests, requiring assessment to be carried out when the learner or candidate is engaged in an extended act of communication;
2. they pay attention to the social roles candidates are likely to assume in the real world settings, and offer a means of specifying the demands of such roles in detail.

The communicative approach to spoken language testing involves assessment of how language is used in real communication. Accordingly, Heaton (1988) states that most communicative language tests aim to 'incorporate tasks which approximate as closely as possible to those facing the students in real life'. Success in actual language performance is judged in terms of the effectiveness of the communication which takes place rather than formal linguistic accuracy. Consequently, the assessment of learners' production of spoken language or test performance should relatively concentrate more on interaction efficacy than on accuracy of language forms.

In addition, the four following characteristics of communicative language tests mentioned by Brown and Gonzo (1995, p.421-422) include a broad basis for both the design and use of language tests.

First, such tests create an 'information gap,' requiring test takers to process complementary information through the use of multiple sources of input. .... The second characteristic is that of task dependency,with tasks in one section of the test building upon the content of earlier sections .... Third, communicative tests can be characterized by their integration of test tasks and content within a given domain of discourse. Finally, communicative tests attempt to measure a much broader range of language abilities – including knowledge of cohesion, functions, and sociolinguistic appropriateness – than did earlier tests, which tended to focus on the formal aspects of language – grammar, vocabulary, and pronunciation.

To put it narrowly for oral testing, all speaking tests that encompass the same purpose of measuring test takers' speaking ability in real interactions are expected to be used to assess authentic language use in context and the ability to communicate meaning, that is to include all the characteristics mentioned above. As previously discussed, the ability to communicate meaning is assured by success of meaning negotiation (figure 2.3 above) in actual acts of interaction.

Speaking tests aim at eliciting test takers' ability of communicating ideas, and how to do this depends upon the content of test tasks or questions that fit students' level of language proficiency. Test takers' different levels of language proficiency can be reflected in the difficulty degree of test tasks. As reviewed in the previous section 2.1, this degree of task difficulty called communicative stress should be taken in account in the teaching and testing of speaking skill, especially on the part of teachers or test developers and assessors. A thorough understanding of the issue helps testers to make informed judgements of 'what type of speaking activity the student would find reasonably 'unstressful' at a particular point in his course' (Brown and Yule, 1983, p.107). Obviously, tasks of oral testing are to be graded mainly in accordance with the degree of communicative stress.

To sum up, the adequate approach, in my viewpoint, to assessing learners' production of spoken language is to measure the extent to which they are able to successfully convey and achieve the intended purposes of a particular test task. In other words, learners' performance on an oral test task should be examined in terms of communicative effectiveness or success of meaning negotiation. However, this assessment way, if a real success, is greatly related to the communicative stress under which test tasks are designed. Therefore, the next two sections will closely review more factors that must be taken into account during the construction process of speaking tests.

**2.3 Theoretical Framework for Oral Test Development**

The two previous sections have discussed the major features of spoken language taken into consideration in assessing production of spoken language, and have considered the communicative approach as the most adequate one to assessing spoken language production. This section describes in detail the theoretical framework for developing language tests which is intended for the following interpretation into the development of speaking tests.

Whether a test is useful or not much depends on test development process. Bachman and Palmer (1996) divide test development into three stages such as design stage, operationalization stage and administration stage.

This process of test development is illustrated in the figure 2.4

Test Development

Design stageAdministration stageOperationalization stage

-Purpose(s) of the test-Tasks in the TLU domain-Characteristics of the test takers-Construct to be measured-Plan for evaluation of test usefulness-Resources- Test task specifications- Blueprint-Giving the test-Collecting test results -Analyzing the results

Figure 2.4: The Model of Test Development

The design stage involves describing and identifying all the factors related to the test. These factors are the purpose(s) of the test as a whole, target language use tasks, test takers' characteristics, language ability to be measured, usefulness of the test and resources.

The operationalization stage consists of two sub-stages. The former is the development of test task specifications referring to the purpose of individual test tasks, the construct to be measured, the setting, time allotment, instructions for responding to the task, characteristics of test input, and scoring method. The latter is the development of a blueprint – a description of 'how test tasks will be organized to form actual tests'. The blueprint is therefore the structure of a test including the number of test tasks/parts and the relative importance of tasks/parts intended for the purpose(s) of the whole test, and the specifications of each test task.

The administration stage involves giving the test to a group of specific test takers, collecting test results, and analyzing the results.

**2.3.1 Design Stage**

The Design stage involves six activities all aiming at producing a design statement as a principled basis for the other two stages (Bachman & Palmer, 1996, p. 88). The six activities are as follows:

(1) Description of the test purpose(s). First specific inferences about language ability and capacity for language use made from the test takers' performance are explicitly stated, and then specific decisions based on these inferences are provided.

(2) Identification and description of test tasks in the target language use (TLU) domain. A set of the TLU task types is characterized as the basis for developing actual test tasks.

(3) Description of the characteristics of the test takers. The characteristics to be believed to be particularly relevant to test development involve personal characteristics, topical knowledge, general level and profile of language ability, and predictions about test takers' potential affective responses to the test.

(4) Definition of the construct/ability to be measured. The components of language ability to be assessed through the test task(s) are critically determined.

(5) Development of a plan for evaluating test usefulness. This plan consists of three parts as follows:

an initial consideration of the appropriate balance among the six qualities of usefulness and the setting of minimum acceptable levels for each,

the logical evaluation of usefulness, and

procedures for collecting qualitative and quantitative evidence during the administration stage.

(Bachman & Palmer, 1996, p.133-134)

(6) Identification of resources and development of a plan for their allocation and management. The resources refer to the people, material and time involved in test development. The balance between the resources available and required for test development should be taken into account in order to provide a good plan for how to allocate and manage them.

### 2.3.2 Operationalization Stage

The Operationalization stage need to be closely examined with the purpose of helping the concerned staff of TNU to equip themselves with a thorough understanding of the stage and then to gradually improve their practices of oral testing.

As mentioned above, this stage focuses on the structure of a test – the blueprint involving the number of test tasks/parts and specifications of each test task.

Test task specifications are described as follows:

1. The purpose of the test task.

2. The definition of the construct to be measured.

3. The characteristics of the setting of the test task

4. Time allotment.

5. Instructions for responding to the task.

6. Characteristics of input, response, and relationship between input and response.

7. Scoring method.

(Bachman & Palmer, 1996, p.172-173)

These test task specifications can be interpreted in the context of oral testing as:

1. The purpose of the test task.
2. Specified components of oral ability to be tested.
3. The place where the task or language act occurs.
4. Expected duration of task performance.
5. Specified and understandable instructions.
6. Areas of linguistic, pragmatic and topical knowledge adequate.
7. Marking key

In order to well operationalize or produce test task specifications, test writers should make informed judgements of major considerations in oral test operationalization reviewed in detail in Section 2.4

### 2.3.3 Administration Stage

The administration of a test mainly involves three activities such as giving the test to a particular group of test takers, gathering test results and analyzing the results (Bachman & Palmer, 1996, p. 91). In particular, procedures for administering a test include preparing the testing environment, communicating the instructions, maintaining a supportive test taking environment, and collecting the test papers. These all aim at 'guiding test takers through the process of taking the test in accordance with the procedures specified in the test blueprint.'

## 2.4 Major Considerations in Operationalization of Speaking Tests

The communicative approach is considered as one of the most adequate way to measure learners' oral language ability. In order to successfully apply this assessment approach, oral test tasks intended for specific tests must be designed in terms of difficulty degree or communicative stress fitting test takers' language proficiency level. Operationalizing a speaking test that fits the test takers' level of language proficiency, test writers or teachers must (1) know the exact level of the test takers, then (2) choose suitable oral test types or elicitation techniques for test tasks, and finally (3) design the method of marking each task. The following three sub-sections discuss these factors, which should be taken into sound consideration during the operationalization of speaking tests.

### 2.4.1 Level Scale

The explicit classification of test takers' language knowledge levels helps to grade test tasks according to the communicative stress. It is displayed on a formal document including established 'criterion levels of oral language proficiency based on the goals and objectives of classroom instruction' (O'Malley & Pierce, 1996, p.65). This document, called a level scale or rating scale by Underhill (1987), is a series of short descriptions of different levels of language ability in terms of test takers' or students' language knowledge. It describes in brief what a typical learner at each level can do so that teachers and assessors can analytically select or grade test tasks that best fit each level, and can easily decide on the score to give each student in a test.

The following is an example of a level scale with four major levels (Table 2.1 on page 17) based on the level scale introduced in 'Hệ thống Định chuẩn Trình độ Ngoại ngữ' của Hội đồng Châu Âu by Vũ Thị Phương Anh and Nguyễn Thị Kim Thư (2003).

| Elementary | Pre-intermediate | Intermediate | Upper-intermediate |
|---|---|---|---|
| -introduce oneself and others.-ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has.-interact in a simple way provided that the other person talks slowly and clearly and is prepared to help.-use very simple expressions related to areas of the most immediate relevance (e.g. very basic family information, | -communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters.-dsecribe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.-simply talk about familiar matters | -use the language in most situations likely to arise when travelling in an area where the language is spoken.-make a simple connected presentation on topics which are familiar or of personal interest.-describe experiences and events, dreams, hopes and ambitions.-briefly give reasons and | -interact with a degree of fluency and spontaneity that makes regular interaction with native speakers without strain for either party.-make a clear and detailed presentation on a wide range of subjects.-give opinions on topical issues and explain the advantages and |

| shopping, local geography, employment). | regularly encountered in work, school, leisure, etc. | explanations for opinions and plans. | disadvantages of various options. |
|---|---|---|---|

Table 2.1: Level Scale of Language Proficiency Based on the Global Scale by Council of Europe

**2.4.2 Oral Test Types and Elicitation Techniques**

When test takers' proficiency level is explicitly identified on the level scale, adequate oral test types and proper elicitation techniques must be critically selected to fit the test takers' level and the testing situation. This sub-section reviews types of oral test in combination with elicitation techniques of this kind of testing, for these two aspects have an interrelated relationship. An elicitation technique involves the procedures of performance for each test task, and a test task itself represents a test type.

Underhill (1987) classifies oral tests/test tasks into four main types: (1) the direct interview type, (2) the pre-arranged information gap tests, (3) tests where the learner prepares in advance, and (4) mechanical/entirely predictable tests. Each type requires some specific techniques to elicit test takers' language performance named elicitation techniques. The four following sub-sections respectively summarize these four oral test types in combination with the involved elicitation techniques

**2.4.2.1 The Direct Interview Type**

The direct interview is the most common and authentic type of oral test; there is no script and no preparation on the test taker's part. The assessor or interviewer, of course, has quite a careful preparation, but not so rigid as to control exactly what the test taker says. This may result in difficulty in assessing the test performance consistently and reliably. (Underhill, 1987, p. 31)

The assessors should be flexible in choosing suitable and feasible techniques to well elicit the task of this type in a specific testing situation. The most common elicitation techniques used in this case are discussion/conversation, interview, form-filling and question and answer.

Discussion/conversation is associated with interaction between two or more people in which the assessor should create the right atmosphere in a very short time so that the test taker can respond to it. The topics discussed and the directions taken by the conversation are the result of this interaction. (Underhill, 1987, p. 45)

Interview, to some extent, is quite similar to discussion/conversation, but an interview is structured. That is to say, the assessor or interviewer maintains firm control and keeps the initiative; whatever the test taker says is in more or less response to the interviewer's questions or statements. (Underhill, 1987, p. 54-56)

Form-filling is a technique in which the test taker and interviewer work together to fill in a form or questionaire. The questions is usually related to the test taker's personal details, professional situation or language needs. Question and answer refers to a set of disconnected questions raised by the tester. The questions are graded according to difficulty to elicit the test taker's opinions on certain topics. This technique may involve using different question types, giving cues for question formation, and naming. (Underhill, 1987, p. 58-59)

**2.4.2.2 The pre-arranged Information Gap Tests**

In such tests, an information gap between two test takers, or between a test taker and the assessor, is deliberately created by the test designer. The test taker's success and speed in bridging that gap are taken as an indication of his oral proficiency. (Underhill, 1987, p. 32)

The elicitation techniques proposed for this type of oral test are learner-learner description and re-creation, picture story and role-play.

Learner-learner description and re-creation technique requires one test taker to describe a design or construction of model building materials to another test taker who has to reconstruct the model from the description alone, without seeing the original. The technique consists of reporting description to partner, map-reading, and comparing models. (Underhill, 1987, p. 56-58)

Picture story is widely used with more advantages than disadvantages. Before performing the test task, the test taker is given a picture or a sequence of pictures to look at. Then the test taker describes the picture(s) or story freely before being asked questions related to the story. This technique includes using several similar pictures, ordering pictures to create a picture story, using live action, and vocabulary naming from pictures. (Underhill, 1987, p. 66-69)

Role-play technique involves two people, each of whom takes on a particular role in a given particular situation. A few minutes just before the test the test taker(s) is given a set of written instructions to get prepared, and then he carries out his role in the given situation. This technique can be used between an assessor and a student, and between students. (Underhill, 1987, p. 51-52)

**2.4.2.3 Tests Where the Learner Prepares in Advance**

Tests of this type give the test taker a sufficient amount time to prepare the task. The preparation time will range from a few minutes for a blank dialogue to several hours or days for a presentation. (Underhill, 1987, p. 33)

The underlying techniques may be oral report, reading blank dialogue, and re-telling a story.

Oral report technique requires the learner to give an oral presentation on a given topic lasting from five to ten minutes. He or she can refer to the notes, but reading aloud is strongly discouraged. The use of such aids as an overhead projector, a board or flipchart diagrams is encouraged if appropriate. At the end of the presentation, the test taker has to answer all the questions raised by the tester. This technique can be applied by making a mini-presentation with limited preparation time, and by identifying a topic of personal interest at a previous stage. (Underhill, 1987, p. 47-49)

Reading blank dialogue is used in the test context in which the learner is provided a dialogue with only one part written in and prepares the missing lines in a few minutes. The interviewer reads through the given lines and the test taker fills in the blanks aloud. (Underhill, 1987, p. 64-66)

Re-telling a story technique requires the test taker to re-tell a story in his own words after reading it. The test taker is not allowed to refer back to the written text once he has begun to re-tell it. This can be carried out by using notes, using a set text, and using an unseen text. (Underhill, 1987, p. 73-75)

**2.4.2.4 Mechanical/Entirely Predictable Tests**

Mechanical-type tests determine in advance what the test taker is expected to say, for there is always a single correct answer. This complete predictability makes such tests unauthentic and non-communicative. Hence, they cannot be used to measure the test taker's oral fluency but to measure grammatical knowledge or the mechanical aspects of speech such as pronunciation, stress and intonation patterns. (Underhill, 1987, p. 33)

Tests of this type encompass such elicitation techniques as reading aloud, sentence transformation, sentence repetition, translating/interpreting, sentence completion, and sentence correction.

Reading aloud technique requires the test taker to read aloud to the tester, either a passage of text, or part of a dialogue in which the tester or another testee reads the other part. This technique may consist of reading scripted dialogue with someone else reading the other part, reading text with phonetic markers, reading sentences containing minimal pairs, spelling aloud, and reading from a table. (Underhill, 1987, p. 76-78)

Sentence transformation is the technique in which the test taker is given a stimulus sentence and is asked to orally transform it into a different grammatical pattern. This technique allows rapid testing of particular structural areas

and an estimation of the test taker's ability to correct himself. (Underhill, 1987, p. 84-85)

Sentence repetition technique is used in a test in which the test taker listens to a set of sentences or utterances, and then repeats them as accurately as possible. The technique may include repeating sentences of increasing length and repetition of sentences with specific language areas. (Underhill, 1987, p. 86-87)

Translating/interpreting technique involves the test taker's target language translation of a short passage of a native-language familiar text. This technique may have such variations as translating in both directions, translating an unprepared passage, translating test in the language laboratory, and translating disconnected words or phrases. (Underhill, 1987, p. 79-81)

Sentence completion technique is associated with test context in which the test taker is asked to complete a series of sentences with the last few words missing from each. The technique may consist of using written tests, using gapfill to check discourse reference, text completion, using spoken cues, and completing a well-known saying. (Underhill, 1987, p. 81-83)

Sentence correction technique presents the test taker with a sentence containing an error. The test taker's task is to identify the error and to correct it. The test taker can also be given a chance to correct his own errors. (Underhill, 1987, p. 84)

The four types of oral test/test task combined with different elicitation techniques are summarized in table 2.3 below.

| Test types | the direct interview type | The pre-arranged information gap | Tests where the learner prepares in advance | Mechanical/entirely predictable tests |
|---|---|---|---|---|
| Elicitation techniques | -Discussion/conversation-Interview-Form-filling-Question and answer | -Learner-learner description and re-creation-Picture story-Role-play | -Oral report-Reading blank dialogue-Re-telling a story | -Reading aloud-Sentence transformation-Sentence repetition-Translating or interpreting-Sentence completion-Sentence correction |

Table 2.3: Oral Test Types and Elicitation Techniques

No test type as well as no single elicitation technique is said to be the best for an oral test task or an oral test as a whole, for each of them has its own advantages and disadvantages. One elicitation technique may be suitable in a testing situation, but inappropriate in other ones. For example, reading aloud technique may be well used to measure elementary learners' pronunciation, intonation and stress, but may be improperly used to measure intermediate learners' speaking ability as this technique is considered to be uncommunicative. Therefore, it is advisable to combine various test types and elicitation techniques in a test of overall oral ability (Underhill, 1987, p. 37-38). This combination depends on the test purpose, and the areas of language competence and ability that intend to be seen in test takers' performance.

Additionally, oral test tasks 'differ with regard to whether they call for the use of static relationships, dynamic relationships, or abstract relationships' (O'Malley & Pierce, 1996, p. 76). These relationships are mentioned in Section 2.1. The selection of oral test types for test tasks is therefore necessarily related to the difficulty degree corresponding to these relationships. Consequently, O'Malley & Pierce (1996, p. 69) propose that the test tasks selected and designed can challenge the language proficiency level(s) of test takers without frustrating them.

**2.4.3 Marking Key**

During the operationalization process of oral tests, the classification of test takers' level of language proficiency is to be carefully considered with the purpose of choosing adequate test types and elicitation techniques in which the test takers' language performance can be best shown. While designing particular test task(s), test writers or teachers should consider and decide how to mark each test task, and therefore build up a guideline helping the assessors to mark each task. This guidline is called a marking key or marking protocol by Underhill (1987).

A marking key is a set of procedures specified in advance that tells assessors what they are supposed to do step by step in the process of marking each test task/question. Test writers can make the marking quicker and more reliable by drawing up a detailed marking guide that tells the marker how to mark each question.

Underhill (1987, p. 95) identifies the aims of a marking key as follows:

- To anticipate problems that the marker is likely to face, and to suggest how to cope with them.

- To maintain the aims of the test by directing the marker's attention to the language areas that are most important, and by giving general guidelines for dealing with unusual responses.

- To describe the purpose of each question/task.

A marking key revealing such aims thus surely helps to increase the consistency of measurement, that is reliability (See 2.5.2). In fact, oral tests are a kind calling for subjective judgement on the part of assessors, and thus do not have as a high degree of reliability as those that require objective judgement such as multiple-choice or cloze tests with either completely right or completely wrong answers. In order to help assessors achieve the highest possible degree of reliability, it is essential to provide them with a comprehensible marking key conveying the three aims identified by Underhill.

The most important factor concerned in a marking key is the distribution of marks to specific speaking sub-skills that are intended to be measured. These speaking sub-skills are named mark categories by Underhill (1987). The kind of categories defined in a test should be based on the teaching program and be cited by the way in which the teaching syllabus expresses the aims of the program (Underhill, 1987). There are two models that mark categories base on: the traditional model of language components and the more recent model of performance criteria. The former refers to the components of language proficiency (grammar, vocabulary, pronunciation and intonation, style and fluency, content, etc.) while the latter mentions the components of language performance or performance criteria (flexibility, accuracy, appropriacy, independence, hesitation, etc.).

The focus on a number of different language sub-skills or categories can also help to improve marker reliability; the assessor is supposed to give each test taker a separate mark for each category. All these separate marks are then combined to give the overall score, which is related to the process of weighting. In most oral tests or test tasks some categories are more emphasized than others according to the test purpose(s), so a weighting system is used as shown in the following example taken from Underhill (1987, p. 97).

Grammarmarked out of 10 then multiplied by 3

Vocabularymarked out of 10 then multiplied by 3

Pronunciationmarked out of 10 then multiplied by 2

Fluencymarked out of 10 then multiplied by 1

Contentmarked out of 10 then multiplied by 1

Total score 10

In sum, it can be asserted that the marking key plays a very essential role in the design of language tests in general, of oral language tests in particular to ensure the quality of reliability. It must be involved in the whole process of test development from the beginning. Language teachers or test developers should thus take the ways to mark test performance into sound consideration throughout the test construction process.

In the oral test operationalization process, consequently, language teachers or test designers must take great care over not only the selection test types and proper elicitation techniques for the intended test tasks but also the design of a marking key for each test task.

## 2.5 Qualities of a Good Test

The previous three sections are concerned with the techniques and procedures for developing oral language tests whereas Section 2.5 is related to the qualities of a good test, i.e. whether the test results can reveal test takers' actual ability to orally use the language. A test used to elicit test takers' actual language proficiency must reveal such qualities as validity, reliability and practicality.

### 2.5.1 Validity

Test validity generally is concerned with the degree to which a test actually measures what it is supposed to measure. In other words, it refers to the correspondence between abilities to be assessed and real indication of these abilities in a test, so a test is said to be invalid when there is no relationship between them. The concept of validity includes such detailed aspects as content validity, construct validity and predictive validity. A test is said to have content validity if its content represents a sample of the language skills, structures, etc. with which it is meant to be concerned ( Hughes, 1989). When embarking on the test construction, a test writer should first draw up a table of test specifications, describing in very clear and precise terms the particular language skills and areas to be included in the test. Not less important is the construct validity of a test. A test with construct validity is capable of measuring certain specific characteristics in accordance with a theory of language behaviour and learning. In other words, construct validity 'examines whether the instrument permit inferences about underlying abilities.' (Cohen, 1996). According to Hughes (1989), the word 'construct' refers to 'any underlying ability or trait which is hypothesised in a theory of language ability'. This ability or trait is defined by Bachman and Palmer (1996) as 'the domain of generalisation to which our score interpretations generalize'. Certain learning theories or constructs are believed to underlie the acquisition of abilities and skills. Another approach to test validity is to measure the degree of the agreement between results of the test and those provided by some important task at some future point.

### 2.5.2 Reliability

If test validity is defined as accuracy of measurement, test reliability is related to consistency of measurement. A reliable test score will be consistent across different characteristics of the testing situation. Unless test scores are relatively consistent, they cannot give any information at all about the ability measured. Another aspect of overall test reliability is rater reliability. Raters must maintain consistency in their own marking standards. This kind of reliability is called intra-marker reliability (Underhill, 1987). Or the same work marked by different raters should produce similar results, which is named inter-marker also by Underhill. If some raters rate more severely than others, the ratings of different raters are not consistent, and the scores obtained could not be considered to be reliable. Oral tests belong to the kind calling for subjective judgement on the part of the marker, so the scores awarded in an oral test cannot be believed to always have such high reliability.

It is also necessary to recognize that inconsistencies cannot be eliminated entirely. Nevertheless, it is possible to minimize the effects of the potential sources of inconsistencies under control in test design (Bachman and Palmer, 1996). Amongst factors affecting test performance, the characteristics of the test tasks are partly under control. In language test design and development, thus, it is possible to minimize variations in the test task characteristics that do not correspond to variations in target language tasks.

Test administration also involved in the concept of reliability has not been given proper attention at some universities at the present time. Administrating a test involves exam invigilators and such test conditions as classrooms, equipment, materials, exam rules and procedures dealing with test takers' cheating.

### 2.5.3 Practicality

Test practicality pertains to 'the ways in which a test will be implemented, and, to a large degree, whether it will be developed and used at all' (Bachman & Palmer, 1996, p. 35). It concerns practical matters such as the amount of time, human and material resources available for constructing a test, administering it, marking it, and interpreting the results. If the test resources required for implementing a test exceed the resources available, the test will be impractical. Human resources are a crucial component of test construction and administration involving such individuals as test writers, scorers or raters, and test administrators as well as clerical and technical support personnel. In fact, not all institutions have sufficient staff to be in charge of all these well-defined roles. One person may be in charge of several functions. Test writers, key personnel in the process of test development, are involved not only in writing tests but also in collecting materials, editing and recording. Material resources include space (the number of classrooms, language labs needed), equipment (typewriters, computers, cassette players, overhead projectors), test materials (test booklets, answer sheets, audiotapes). Time consists of test development time and the time required to complete the parts of each stage of the test development process.

Moreover, the specific types and amounts of resources required may differ according to the design of a specific test, and available resources may vary from one situation to another. Test practicality thus can only be determined for a specific testing situation. Obviously, to determine the practicality of a given test, test developers must take into account of the resources required to develop a test, and the management and allocation of the resources available.

## 2.6 Summary

Chapter 2 has considered the main features of oral testing, particularly spoken language and considerations on how to elicit students' overall speaking ability. Production of spoken language is examined in a continuum of the language functions and a success of meaning negotiation. Next, the communicative approach to assessing production of spoken language is considered one of the best ones. And Bachman & Palmer's theoretical framework for test development is reviewed as the basis for description and evaluation of current oral testing practices at TNU. Also, this framework is used as the main foundation on which suggestions for improvement of TNU oral testing problems are based. In addition, major considerations in operationalizing speaking tests include a level scale, selection of test types, elicitation techniques, and a marking key. Finally, for a test to be valid, it must also be reliable and practical. Validity is associated with accuracy of measurement, and reliability refers to the consistency of measurement. Practicality, more or less important, concerns the ways in which the test will be implemented in a given situation, or whether the test will be used at all.

## CHAPTER 3: methodology

Chapter 2 has reviewed major issues in oral language testing in order to provide an adequate understanding of its theory, which serves as the basis for investigation of the current practices at TNU. In particular, the discussion of Bachman and Palmer's theoretical framework for developing tests presents the basis for the evaluation of oral testing practices at TNU, and for suggestions for improving its drawbacks. In order to investigate TNU current oral testing practices, the researcher (1) analysed the present process of oral test development at this institution, and (2) surveyed the staff's perceptions of oral language testing. This chapter consists of three sections. The two research

questions are first raised. The second sub-section respectively presents the data collection instruments used to carry out (1) & (2). The other sub-section describes the procedure for conducting the study.

## 3.1 Research Questions

The study aims at answering the two following questions.

1. What are strengths and weaknesses of the oral test development procedure at TNU?

2. What are the English teaching staff's perceptions of oral language testing?

## 3.2 Data Collection Instruments

In order to answer the two research questions, the researcher collected information from two sources: (1) a situation analysis and (2) a questionnaire. Firstly, the situation analysis was carried out with a checklist based on Bachman and Palmer's framework for test development. To ensure the reliability of the information from the situation analysis, one end-of-term speaking test was observed and tape-recorded. Secondly, the questionnaire was developed to elicit the staff's perceptions of oral language testing. Therefore, this sub-section respectively describes these three instruments

### 3.2.1 The Checklist

The situation analysis with a checklist (as seen on page 31) that has been developed based on Bachman and Palmer's framework for test development reviewed in 2.3 involved such four factors as (1) the test design stage, (2) the test operationalization stage, (3) the test administration stage, and (4) the use of test results.

| 1. Test Design Stage | |
|---|---|
| Are the purposes of oral tests explicitly identified? | |
| Which kind(s) do the oral tests include? Selection PlacementDiagnosisAchievement | |
| Is a set of the TLU tasks presented? | |
| Is there an official document including detailed instructions on students' language proficiency levels? | |
| Is there an official document including detailed instructions on construct or language ability to be measured? | |
| Are there any criteria set for evaluating test quality? | |
| 2. Test Operationalization Stage | |
| Are there any official guidelines on the number of test tasks to be included in a particular speaking test? | |
| Are the specifications of each test task provided? | |

| | |
|---|---|
| 3. Test Administration | |
| Are the assessors informed of how to mark the test tasks before the test is administered? | |
| Is the testing environment well prepared? | |
| Is a supportive test taking environment maintained? | |
| Are the instructions for each test task made clear to the students? | |
| 4. Use of Test ResultsIs the information from test results used for | |
| ...grading the students? | |
| ...evaluation of the effectiveness of instructional programmes? | |
| ...the teachers' modification of teaching methods and materials? | |

**3.2.2 The Observation**

Observation of a particular achievement speaking test is expected to help collect evidence that supplements the analysis of the current oral test development process, namely the operationalization and administration stages. The observation enables the researcher to directly collect data firsthand, and the data gathered describes the observed phenomena as they take place in their natural settings (Nachmias, 1996, p.206); therefore, this kind of information is surely of great reliability.

The observation focused on the oral test type(s) with underlying elicitation techniques being used (See Appendix 2), time spent on the students' test performance, and interaction between the assessors and test takers recorded in a tape and then transcribed (See Appendix 3).

**3.1.3 The Questionnaire**

In order to elicit TNU staff's perceptions of oral language testing, questionnaires were formulated, and then delivered to the staff members. Questionnaires, in this thesis, are chosen as an adequate way to well elicit respondents' knowledge, for the respondents are not put under pressure of time, i.e. they answer questions in their own time and at their own pace, and in an anonymous style of responding they undoubtedly feel free and comfortable to answer questions (Gillham, 2000; Nachmias, 1996).

The respondents include 12 English language teachers of TNU, who are involved in the development process of oral tests. They are at the age of 27 to 50. They all had tertiary training in language teaching in different educational institutions throughout Vietnam.

The questionnaire consists of 10 questions, 8 of which are used to elicit the teaching staff's perceptions of oral testing. These 8 questions are developed based on the theory in oral language testing reviewed in Chapter 2. The other two concern the staff's working experience, and their qualifications in language testing. The questions are as follows:

Questions 1 and 2 relate to the functions of spoken language and the communicative approach to measuring speaking skill.

Questions 3 and 4 relate to oral test types and elicitation techniques used in a test of speaking ability.

Question 5 relates to grading test tasks and validity.

Question 6 relates to the design and operationalization process of oral tests, and reliability.

Questions 7 and 8 relate to reliability of a test.

In general, questions 1 to 8 are close questions, and questions 9 & 10 are open-ended.

For questions 1 and 2, the respondents are required to arrange the options according to their level of priority. Question 3 requires 1 out of 4 options. Questions 4 and 5 require the respondents to tick the level(s) of language proficiency that the elicitation techniques and particular questions fit. Question 6 and question 8 allow more than one choice. For question 7, the respondents are required to choose 1 out of 3 options.

Questions 9 and 10 are meant to investigate possible sources of the differences in the subjects' perceptions of oral language testing.

## 3.3 Procedures

The study meant to investigate TNU current oral language testing practices involves two steps. Firstly, the reseacher analysed the development process of speaking tests at this institution from School Year 1998-1999 (,when the researcher started working at this institution,) up to now. The current practices are evaluated in order to point out the strengths and weaknesses based on Bachman and Palmer's theoretical framework for test development reviewed in 2.3 (Chapter 2).

To ensure the reliability of the information from the analysis, one end-of-term speaking test was observed and tape-recorded. The information from this observation is particularly meant to confirm the evaluation of the test operationalization and administration stages. The test under observation was used for the second-year students at the end of Term 2 - School Year 2002-2003 (See Appendix 2). This test was administered in the morning on June 25th, 2003 at TNU. 10 students were chosen at random to be audio-recorded. Their performance was recorded in a tape and then transcribed (See the tapescript in Appendix 3).

There were 60 test takers altogether, and they were divided into 3 groups of 20 in 3 separate rooms. However, because of the time limitation and the simultaneity of the 3 groups, only 10 of them were chosen at random for recording. Each group was conducted by 2 assessors. The students drew lots for the test task or topic to prepare for 5 to 10 minutes before performing it. The test consisted of 8 topics altogether (See Appendix 2). Both the assessors and the students were unaware of being recorded.

Secondly, the investigation of TNU current oral testing also involved a survey of the English teaching staff's perceptions of oral skill assessment. The survey was conducted in the form of questionnaires. The questionnaires were delivered to the respondents and collected one week later. The respondents were clearly informed of the purpose of the questionnaire.

The questionnaire (See Appendix 4) is written in Vietnamese to make sure that the respondents' different extent of familiarity of some technical terms in language testing does not affect their understanding of the questions and thus distort their responses.

## 3.4 Summary

Chapter 3 has raised the two research questions and described the research methods employed, particularly the subjects of the study, the data collection instruments used to serve the purposes of the study, and the way the study was conducted. The data collection instruments include a checklist for oral test development summarizing the oral test development process at TNU described, tape recordings of actual test performance used to supplement the operationalization and administration processes of achievement speaking tests, and questionnaires formulated to investigate the staff's perceptions of oral testing. The data gathered from the study will be thoroughly analysed in the next chapter.

# CHAPTER 4: RESULTS AND DISCUSSION

Chapter 3 has outlined the two research methods used to carry out the study: the situation analysis and the questionnaire survey. The situation analysis is meant to evaluate the current oral test development at TNU while the survey is aimed at investigating the staff's understanding of oral language testing. Chapter 4 is thus divided into two sections evaluating (1) the current development process of speaking tests and (2) the staff's perceptions of oral skill assessment.

## 4.1 Evaluation of TNU Current Development Process of Oral Language Tests

As previously mentioned, the analysis of TNU current development process of speaking tests and the observation of one real end-of-term speaking test are intended for the evaluation of the practices of developing speaking tests. This section (1) starts with a detailed review of the present development process of speaking tests summarized in the checklist, (2) presents the results gathered from the observation of the particular achievement speaking test, and (3) analyses the results in order to reach a conclusion of whether TNU current procedures for oral test development are consistent with the theoretical framework or not.

### 4.1.1 Review of TNU Current Development Process of Oral Language Tests

Information regarding the existing practices of developing English speaking tests at TNU described in this sub-section plays an essential role in the critical evaluation of these practices. TNU current oral test development process is described in relation to such four main factors as (1) the test design stage, (2) the test operationalization stage, (3) the test administration stage, and (4) the use of test results.

First of all, the oral tests used at this institution have been formally administered at the end of each term in order to measure what the students have actually achieved after one particular time of learning. Such kind of test is called a final achievement test by Hughes (1989) and McNamara (2000), progress and grading test by Bachman & Palmer (1996), course test by Davies (2000), or final or attainment test by Heaton (1990). All the teachers or test writers as well as assessors have always known that they should elicit the students' actual ability to use the language in real communication, and especially their language knowledge and ability they are required to grasp by the end of one particular term. Therefore, oral tests at this institution are explicitly identified as achievement ones from the very start, and both the teachers and the assessors have been aware of this.

However, neither the Department nor the English Section has produced a formal document including explicit classification of students' language proficiency levels, particular areas of language ability or construct to be assessed and sets of TLU tasks identified for all the levels. Also, they have not established any criteria for test quality evaluation.

Secondly, as regards the operationalization process, the test designers have not kept in hand the official document mentioned above with descriptions of different levels of language proficiency in terms of the students' language knowledge, i.e. a level scale (discussed in 2.3.1) and with areas of language ability to be tested. Additionally, the teachers or test writers have not received any detailed guidelines or instructions on number of test tasks included in the test(s) as well as test task specifications. Nevertheless, they have been informed about the administration time of the test(s) in advance in order to ensure punctual test production and submission.

Therefore, the teachers or test designers have freely produced the speaking tests in their own way, and most of the oral tests conducted make use of merely one oral test type – Tests where the learner prepares in advance – and one elicitation technique – oral report – and consist of one test task/part (See Appendix 1- these three tests were used for the same class for three terms in succession). Furthermore, as shown in these three tests, none of the test task/question is attached with neither external prompts helping the students make a structured presentation nor explicit instructions quantifying language knowledge and ability needed to perform the task.

Thirdly, concerning the administration process, there has been no detailed guidance in the form of either a meeting among the group of administrators and assessors or an official document regarding the students' language

proficiency level, i.e. level scale mentioned above, and no guidelines on method(s) of marking students' performance on each test task. In short, the assessors are never informed of or provided with these two important things before test administration.

At the end of one term the oral test administration takes place with three classes of 60 students on average. The test administration for one class is allotted half a day and each class is usually divided into two groups of about 30 students in two separate rooms. The time for test performance of each student is about 4 to 5 minutes, hence. During test administration every 5 students are called into to draw lots for test questions or tasks to make a preparation for 5 to 10 minutes. Then each student presents his/her preparation in front of two assessors. The two assessors often raise some questions related or perhaps unrelated to the student's presentation. Sometimes the assessors do not ask any questions. The student's final score is taken from the average of marks given by the two assessors. Meanwhile, the other students waiting for their turn are standing along the corridor and talking, that is to say a supportive testing environment is not maintained.

Last but not least, after all the students finish their performance, test results are analysed to grade the students. However, the teachers are not provided with and are not allowed to keep the list of students' test scores. The oral test results at this institution are not thus used to either evaluate the effectiveness of the instructional programs or modify the teaching methods and materials.

TNU oral test development described above can be reviewed by means of the checklist (Table 4.1 on page 39), specially designed with the purpose of highlighting the strong and weak points of the current practices. The answer 'Yes' is ticked ' ' and 'No' is crossed '×'.

| 1. Test Design Stage | |
| --- | --- |
| Are the purposes of oral tests explicitly identified? | |
| Which kind(s) do the oral tests include? Selection PlacementDiagnosisAchievement | |
| Is a set of the TLU tasks presented? | × |
| Is there an official document including detailed instructions on students' language proficiency levels? | × |
| Is there an official document including detailed instructions on construct or language ability to be measured? | × |
| Are there any criteria set for evaluating test quality? | × |
| 2. Test Operationalization Stage | |
| Are there any official guidelines on the number of test tasks to be included in a particular speaking test? | × |
| Are the specifications of each test task provided? | × |
| 3. Test Administration | |
| Are the assessors informed of how to mark the test tasks before the test is administered? | × |
| Is the testing environment well prepared? | × |
| | |

| | |
|---|---|
| Is a supportive test taking environment maintained? | × |
| Are the instructions for each test task made clear to the students? | |
| Are the test tasks in use attached with any limitation of knowledge? | × |
| 4. Use of Test ResultsIs the information from test results used for | |
| ...grading the students? | |
| ...evaluation of the effectiveness of instructional programmes? | × |
| ...the teachers' modification of teaching methods and materials? | × |

Table 4.1: A checklist for Oral Test Development

Table 4.1 partially speaks for TNU current oral testing practices with many '×s', which reveals impropriety in the speaking test development at this institution and undoubtedly indicates a big gap between practice and theory.

**4.1.2 The Observation Results**

Table 4.2 displays oral test types used during the administration of the end-of-term speaking test for the second-year students (Term 2 – School Year 2002-2003 – Appendix 2). The information from this table is intended for evaluation of oral test types in use in the next sub-section 4.1.3

| Students | Oral test types | | | |
|---|---|---|---|---|
| | Direct interview | Pre-arranged information gap | Tests where the learner prepares in advance | Mechanical /entirely predictable tests |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| | | | | |

| | | | | |
|---|---|---|---|---|
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Table 4.2: Summary of Oral Test Types Used in the Achievement Speaking Test for the Second-Year Students (School Year 2002-2003)

Table 4.2 indicates that only one test type was in use, yet as discussed in 2.4.2 (Chapter 2), a speaking test, namely an achievement one intended to measure overall oral proficiency, that can be believed to be valid should be a combination of various oral test types, at least two.

The following is Table 4.3 presenting elicitation technique(s) employed to elicit the 10 students' ability during the achievement test mentioned above, their topic number or test question, duration of their test performance, their interaction with the assessors. All these details were recorded and transcribed in Appendix 3.

| Students | Elicitation techniques involved in tests where the learner prepares in advance | | | Topic number | Time(minutes) | Interaction focus |
|---|---|---|---|---|---|---|
| | Oral report | Reading blank dialogue | Retelling a story | | | |
| 1 | | | | 8 | 5 | Student's presentation without any questions from the assessors |
| 2 | | | | 2 | 4.5 | Student's presentation without any questions from the assessors |
| 3 | | | | 1 | 5 | Student's presentation with 1 question raised by the assessor |

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | | | | 8 | 3 | Student's presentation without any questions from the assessors |
| 5 | | | | 6 | 3 | Student's presentation without any questions from the assessors |
| 6 | | | | 2 | 2.5 | Student's presentation without any questions from the assessors |
| 7 | | | | 3 | 3.5 | Student's presentation without any questions from the assessors |
| 8 | | | | 5 | 2.5 | Student's presentation without any questions from the assessors |
| 9 | | | | 1 | 5 | Student's presentation with 2 questions raised by the assessor |
| 10 | | | | 4 | 3 | Student's presentation without any questions from the assessors |

Table 4.3: Summary of the Students' Oral Test Performance in the Achievement Speaking Test for the Second-Year Students

As can be seen in table 4.3, oral report, one of the three main elicitation techniques used to elicit test takers' speaking ability through their performance on this kind of test – Tests where the learner prepares in advance (See 2.4.2.3, Chapter 2), was the only elicitation technique employed throughout this achievement test. Additionally, after most of the students finished their presentation, the assessors did not raise any questions except for students 3 and 9.

**4.1.3 Analysis of the Results**

The evaluation of TNU current oral testing practices is carried out in relation to four factors described in 4.1.1: (1) test design stage, (2) test operationalization stage, (3) test administration stage, and (4) use of test results.

- Test Design Stage

As can be easily seen in Table 4.1, oral tests are explicitly identified as achievement ones from the very start. Obviously, clear identification of test type at the beginning of a course proves to be beneficial because the teachers can integrate the test content into the teaching program. As pointed out by Brown (1994), Heaton (1990), Hughes (1989) and Ur (1996), achievement tests should be integrated into the teaching program and related directly to the classroom lessons or units, the syllabus or curriculum. Therefore, information or indication of students' performance on an achievement test reveals their achievement or progress at the end of a course of study (Bachman & Palmer, 1996), and an achievement test of speaking skill is of course a means of eliciting students' progress in overall speaking ability after a course/term of study.

However, a product of this stage involving such four crucial things as students' profile of language ability, construct/ability to be measured, sets of test tasks in the TLU domain and a plan for test quality evaluation, as described in 4.1.1, has never been produced and presented to the teachers as a principled basis or guidelines for the other two stages. This undoubtedly indicates that the first stage of oral test development at TNU is far from being consistent with the theoretical framework reviewed in 2.3.1 – Chapter 2. As a result, this big mismatch leads to the staff's improper practices in the other two stages.

- Test Operationalization Process

Apart from the mismatch between practice and theory at this institution mentioned above, a remarkably essential fact shown in Table 4.1 is that the Department and English Section have not provided any specific guidance, i.e. a blueprint, for speaking test construction process, namely (1) the number of test tasks to be included in a speaking test, and (2) specifications of each test task. These two factors are critically analysed respectively.

Firstly, as previously discussed, an achievement test of speaking skill is a means of eliciting students' progress in overall speaking ability after a course of study, yet most of the achievement speaking tests in use at TNU can be asserted to fail to serve this purpose because they make use of merely one type of oral test or one test task – Tests where the learner prepares in advance (Tables 4.2) - combined with only one elicitation technique - Oral Report (Table 4.3). This is partially because no blueprint is presented. Underhill (1987) points out that an oral test rarely consists of only one elicitation technique but it is usual that it involves several techniques placed in a sequence. The reasons he provides for including more than one technique in an oral test are as follows

1. It is more authentic to use a mix of techniques, with the learner doing different things with the language....
2. An oral test that consists only of Question and Answer, for example, will naturally favour learners who are good at answering questions....
3. To help improve the consistency of assessment, a change of tasks during a test can be used as an opportunity to swap interviewers and so combine multiple tasks with multiple assessment....
4. A live test with several different parts is more flexible and can be adapted quickly to meet changing circumstances or different needs....

(Underhill, 1987, p.38)

Probably, such test tasks have been carefully discussed in class, and the students are expected to produce 'well-prepared' talk, even predictable questions can also be prepared in advance. Of course, 'the task(s) on which the student has to perform may be generally familiar in form to the student, but the student cannot 'prepare' a written version of what he will say' (Brown & Yule, 1983, p.120). He must prove to the assessors that in his test performance he has learned to use, not to repeat, what he has been taught. What we as examiners want to know when testing a students is not whether the students has learned what to have been taught, but whether he is able to produce an extended piece of spoken English appropriate to the communicative situation he encounters (Brown & Yule, 1983, p.120).

Obviously, this popular kind of oral test at TNU is far from being useful in measuring the students' overall language oral proficiency, and can be said to be lacking in construct validity and reliability (See 2.5, Chapter 2).

Secondly, no specifications of particular test task(s), especially specified components of oral ability to be tested, areas of language knowledge adequate and a marking key, to some extent, results in the teachers' or test designers' inadequate and useless tests. It can be said that there is lack of consideration of communicative stress in the oral test construction.

As can be seen in four achievement speaking tests (See Appendices 1 & 2), all the test questions/tasks – topics- are never accompanied with any external prompts helping the students make a structured presentation, and any explicit instructions quantifying language knowledge and ability needed to perform the tasks.

It is extremely necessary for test writers to provide clear instructions helping test takers to organise a spoken presentation for test performance because students are always encouraged to produce effectively organised speech so that the listener finds it easy to catch up with what is being said (Brown & Yule, 1983, p.119).

Also, in order to write test tasks fitting students' proficiency levels, test writers need really give explicit instructions quantifying language knowledge and ability. The quantification of performance on a particular task much depends on the grading of tasks according to cognitive difficulty (Brown & Yule, 1983, p.121). To put in another way, the same task type can be made easier or more difficult. For example, describing a room with 8 elements is apparently more difficult than a room with 5 elements. Inevitably, test designers or teachers of speaking skill should always bear in mind informed judgements of the degree of this cognitive difficulty or communicative stress (Figure 2.2, Chapter 2) during test operationalization process.

Besides, no official instructions on criteria for marking students' test performance are presented; thus, the test writers/teachers are unaware of the importance of scoring method(s) for each test task, and they never design a marking key (See 2.4.3, Chapter 2) instructing assessors how to assess students' performance on test tasks. As discussed in 2.4.3, in a marking key, language and skill categories are identified and awarded separate marks according to test purpose(s). As Underhill (1987, p.94) points out the aim of a marking key is 'to save time and uncertainty by specifying in advance, as far as possible, how markers should approach the marking of each question or task'. With help of a marking key and a level scale mentioned above, assessors can mark a test more quickly and reliably, for each language or skill category is expected to be separately marked.

- Test Administration Process

Table 4.1 and 4.3 indicate that TNU speaking test administration reveals many a shortcoming. These weak points include (1) lack of test administration standardisation, (2) lack of reliability in marking test takers' test performance, and (3) lack of supportive testing environment.

First, before test administration there has been no official meeting - named 'the standardisation meeting ' by Alderson, Clapham & Wall (1995)- for discussion and agreement on how to mark each question/task among the group of assessors. Perhaps the administrators here tend to think the assessors, as language teachers, must obviously know how to fully elicit the students' oral proficiency, so they do not need to be informed of what to do during the test. Even when the assessors can be aware of the importance of this meeting, they are unable to hold it. It is partially because the staff's insufficient knowledge of oral testing cannot help them to design an appropriate marking key and a reasonable description of mark categories with a mark criterion.

Therefore, before test administration, a marking key and a mark criterion for mark categories are first needed from test designers, and then a considerable amount of time must be spent on discussion to reach agreement on the way to mark each question/task. Alderson, Clapham & Wall (1995, p.112) maintain, 'although this is likely to be expensive, it is the safest way of ensuring that enough discussion will take place for all examiners to understand thoroughly the level scale and the procedures for scoring.' All these things aim at assuring reliability of an achievement speaking test.

Second, Table 4.3 reveals that, during the students' test performance, interaction hardly existed between the assessors and the test takers or students apart from 2 students out of 10. These two students were asked 1 or 2 questions. Moreover, the duration of these 10 students' test performance varies 2 minutes on average. As discussed in 2.1- Chapter 2, spoken language has two functions, interactional and transactional, which are both necessarily incorporated into a speaking test. In fact, in most of the oral tests in use at TNU, namely the achievement test mentioned above, the students are expected to merely produce transactional instances of the language. Can such

tests be considered to be able to measure test takers' or students' overall oral proficiency? The answer is surely no because they reveals no interactive communication. This also means that the assessors gave scores just on the students' presentation, which also surely indicates a lack of validity and reliability (See 2.5, Chapter 2).

Last but not least, as regards a supportive testing environment the oral tests were almost administered in noisy rooms. Students should be put at ease before and during their performance, which can increase their confidence. Bachman & Palmer (1996), hence, demonstrate that it is crucial to maintain a supportive environment throughout the test, that is to avoid distractions due to temperature, noise, excessive movement, etc. In order to do this, test administrators and assessors are to be in control of techniques and create an atmosphere which will help each student to feel at ease (Alderson, Clapham & Wall (1995, p.116). For those students waiting for their turn should be sitting in a comfortable room, not standing along the corridor and talking so as not to affect the others' performance.

- Use of test results

The last factor under evaluation involves ways of how test results or students' final scores in test performance are used. As previously described in 4.1, students' oral test scores are used to grade them in terms of their progress or achievement after a term/course of study. This is the most popular and common purpose of all achievement tests, that is to make the final decision on students' proficiency kept in their study record in the form of grades. Furthermore, teachers and students are really interested in receiving feedback on students' progress which helps students 'guide their own subsequent learning', and helps teachers 'modify their teaching methods and materials so as to make them more appropriate for their students' needs, interests and capabilities' (Bachman & Palmer, 1996, p. 98). However, TNU students' test scores have never been used to either evaluate the effectiveness of instructional programs or make any improvement in teachers' teaching methods and materials. In other words, oral testing at this institution has no effect on the teaching and learning of speaking skill which is named negative washback or backwash by Hughes (1989), Heaton (1988) and McNamara (2000). Information regarding inferences about students' proficiency made from test performance can be really useful for assessing the efficiency of a teaching program as well as teachers (Bachman & Palmer, 1996, p. 98).

In conclusion, the analysis of TNU current practices of developing oral language tests reveals a number of weaknesses as follows:

There is no principled basis for oral test operationalization and administration

Oral tests in use lack construct validity and reliability

There is lack of consideration of communicative stress in oral test operationalization

There is lack of test administration standardisation

There is lack of a supportive test taking environment

These current practices are thus far from being consistent with the theoretical framework for test development.


**4.2 Evaluation of TNU staff's Perceptions of Oral Testing**

As previously mentioned, the questionnaire survey was carried out on 12 members of TNU English teaching staff with the purpose of investigating their perceptions of oral language testing. The questionnaire consists of 10 questions, 8 of which help to elicit the staff's perceptions of oral testing, and the other two indicate the staff's working experience and their qualifications in language testing. Concerning information elicited about the staff's perceptions, in particular, questions 1 & 2 reveal their awareness of the two functions of spoken language, questions 3 of the number of oral test types and elicitation techniques involved in a test, questions 4 & 5 of communicative stress in tasks suitable to students' proficiency levels, and question 6 of the procedure for oral test design and operationalization. Meanwhile questions 7 & 8 elicit the teachers' reliability degree within their own inferences about students' oral test performance. This section (1) presents the data collected from the teachers'

responses to all the questions of the questionnaire, and (2) analyses the data to evaluate their understanding of this kind of assessment.

### 4.2.1 Results

This sub-section first provides the list of the correct answers to the questions and the results collected from the survey.

List of the correct answers to the questions

The correct answers for the questions of the questionnaire are provided in Table 4.4 below.

| Questions | Answers |
| --- | --- |
| 1 | d-f-h-e-b-c-a-g |
| 2 | i-h-f-d-g-e-b-c-a |
| 3 | b, c or d |
| 4 | Year 1: b, c, d, e, h, i Year 2: a, b, c, d, f, g, h Year 3: a, b, c, f |
| 5 | Year 1: b. e, i Year 2: a, d, gYear 3: c, f, h |
| 6 | All the options must be selected |
| 7 | Selected in the respondent's own opinion |
| 8 | Used when option b or c of question 7 has been chosen |

Table 4.4: Correct Answers for the Questions in the Questionnaire

As far as the two functions of spoken language concerned, criteria for assessing test takers' interactional and transactional short turns, as discussed in 2.1 (Chapter 2), are more based on their communicative reaction and successfully negotiated ideas rather than on content, size, cohesion or coherence like in taking transactional long turns. Thus, the correct order for Question 1 is d-f-h-e-b-c-a-g, and the correct order for Question 2 is i-h-f-d-g-e-b-c-a. However, every two positions of the last fourth (Question 1) and of the last fifth (Question 2) positions can be exchanged, but the first fourth positions cannot be put in the last fourth (Question 1) or the last fifth (Question 2) positions. That is to say, the order d-f-h-e of question 1, and i-h-f-d of question 2 cannot be changed, and these options must be given the highest priority based on the major features of communicative language testing, that is, interaction efficacy is the assessment prerequisite not accuracy of language forms.

For Question 3, related the number of test tasks/parts included in a test of overall oral proficiency, it is necessary to make use of at least 2 elicitation techniques representing 2 oral test types as demonstrated in 2.3.2 (Chapter 2). The correct answer to this question, hence, is either b, c, or d.

Concerning grading test tasks according to degree of communicative stress reviewed in 2.1 (Chapter 2), elicitation techniques suitable and adequate for each level of proficiency in Question 4 are as follows: Year 1: b, c, d, e, h, i; Year 2: a, b, c, d, f, g, h; and Year 3: a, b, c, f. Similarly, specific tasks in Question 5 used to measure speaking ability for each level are identified as follows: Year 1: b, e, i; Year 2: a, d, g; and Year 3: c, f, h.

As reviewed earlier in 2.4 (Chapter 2), all the steps, in Question 6, must be involved in the oral test design and operationalization process. Therefore, all the options of this question are to be chosen.

The data collected from the questionnaire

Table 4.5 on page 49 presents information regarding the subjects' own criteria for assessing test takers' performance on interactional and transactional short turns (Question 1).

| OptionsLevel of priority | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 6 | 3 | 1 | 0 | 2 |
| 2 | 0 | 2 | 0 | 4 | 1 | 3 | 0 | 2 |
| 3 | 3 | 2 | 1 | 1 | 2 | 1 | 0 | 2 |
| 4 | 1 | 1 | 6 | 0 | 2 | 0 | 0 | 2 |
| 5 | 1 | 3 | 3 | 0 | 2 | 0 | 1 | 2 |
| 6 | 4 | 2 | 1 | 0 | 1 | 2 | 0 | 2 |
| 7 | 3 | 2 | 0 | 1 | 1 | 5 | 0 | 0 |
| 8 | 0 | 0 | 1 | 0 | 0 | 0 | 11 | 0 |

Table 4.5: Teachers' Assessment Priority Perception of Interactional and Transactional Short Turns

Most of the respondents (11/12), as shown in Table 4.5 give option (g) the lowest priority in their assessment criteria for interactional and transactional short turns.

Table 4.6 displays the respondents' perception of priority in assessing students' test performance on transactional long turns (Question 2).

| OptionsLevel of priority | a | b | c | d | e | f | g | h | i |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 3 | 2 | 0 | 0 | 0 | 6 |
| 2 | 3 | 1 | 0 | 2 | 1 | 2 | 0 | 2 | 1 |
| 3 | 0 | 3 | 1 | 1 | 4 | 1 | 0 | 1 | 1 |
| 4 | 1 | 2 | 5 | 1 | 1 | 0 | 0 | 2 | 0 |
| 5 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 3 | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 3 | 2 | 1 | 0 | 2 | 1 | 0 | 1 | 2 |
| 7 | 3 | 0 | 2 | 4 | 0 | 1 | 0 | 2 | 0 |
| 8 | 1 | 2 | 0 | 0 | 1 | 6 | 0 | 1 | 1 |
| 9 | 0 | 0 | 1 | 0 | 0 | 0 | 11 | 0 | 0 |

Table 4.6: Teachers' Assessment Priority Perception of Transactional Long Turns

Table 4.6 reveals most of the respondents (11/12) also give option (g) the lowest priority in their assessment of transactional long turns.

Table 4.7 displays all responses the respondents have chosen for Question 3, concerned with the number of test tasks/questions included in an achievement speaking test.

| Number of tasks included | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Number of respondents | 4 | 4 | 3 | 1 |

Table 4.7: Teachers' Choice of Number of Tasks for a Speaking Test

Table 4.7 indicates that 8 out of 12 respondents think a speaking test should make use of two or more tasks/parts.

Table 4.8 reveals the subjects' choice of elicitation techniques for their tests of speaking (Question 4). All the elicitation techniques in the questionnaire are popularly used in tests of oral ability. The appropriate elicitation techniques for each proficiency level, as previously asserted in Table 4.4, are provided in the brackets right below each option.

| OptionsLevels | a(2,3) | b(1,2,3) | c(1,2,3) | d(1) | e(1) | f(2,3) | g(2) | h(1,2) | i(1) |
|---|---|---|---|---|---|---|---|---|---|
| Year 1 | 5 | 4 | 6 | 2 | 7 | 3 | 6 | 8 | 11 |
| Year 2 | 5 | 5 | 2 | 8 | 5 | 11 | 5 | 6 | 1 |
| Year 3 | 6 | 11 | 8 | 4 | 0 | 4 | 2 | 2 | 1 |

Table 4.8: Teachers' Choice of Elicitation Techniques for Levels of Proficiency

As shown in table 4.8, only 2 out of 12 respondents choose (d) for year 1 and (c) for year 2, and half of them choose (g) for year 1.

Table 4.9 on page 51 displays the respondents' selection of specific tasks used to measure test takers' speaking ability at each proficiency level (Question 5). The particular tasks adequate to the three levels, as given in Table 4.4, are presented in the brackets next to each option. For example, a(2) – a refers to one option of the question and (2) to the second level of proficiency (Year 2).

| OptionsLevels | a(2) | b(1) | c(3) | d(2) | e(1) | f(3) | g(2) | h(3) | i(1) |
|---|---|---|---|---|---|---|---|---|---|
| Year 1 | 4 | 9 | 2 | 0 | 2 | 1 | 3 | 3 | 9 |
| Year 2 | 6 | 3 | 9 | 3 | 6 | 1 | 6 | 8 | 3 |
| Year 3 | 3 | 1 | 4 | 10 | 8 | 11 | 5 | 2 | 3 |

Table 4.9: Teachers' Choice of Specific Test Tasks for Level of Proficiency

Table 4.9 reveals that only 2 out of 12 respondents use (e) for year 1, 3 use (d) for year 2, and 2 use (h) for year 3. In addition, up to 9 of them use (c) for year 2, and 10 use (d) for year 3.

Table 4.10 presents responses selected for Question 6, associated with the procedure for oral test design and operationalization. The correct answer, as given in Table 4.4 is selection of all the options.

| Options | a | b | c | d | e | f |
|---|---|---|---|---|---|---|
| Number of respondents | 8 | 9 | 10 | 10 | 6 | 7 |

Table 4.10: Teachers' Choice of Steps to Be Considered in Oral Test Design and Operationalization

Table 4.10 shows that only half of the respondents (6/12) consider option (e) to be essential in the construction process of oral tests.

Table 4.11 reveals information regarding the respondents' reliability degree within their own inferences about students' oral test performance (Question 7).

| Marks given on test takers' performance | Sure | Not very sure | Not sure |
|---|---|---|---|
| Number of respondents | 2 | 10 | 0 |

Table 4.11: Teachers' Confidence in Students' Test Results

As shown in table 4.11, 10 out of 12 respondents do not believe that their marks given on the students' oral test performance can reveal the students' actual interactive ability.

Table 4.12 presents responses selected for Question 8, used to elicit the reason(s) why the respondents are not or not very sure of the marks they have given on the students' oral test performance.

| Options | a | b | c | d |
|---|---|---|---|---|
| Number of respondents | 10 | 5 | 7 | 5 |

Table 4.12: Teachers' Lack of Confidence in Students' Test Results

A remarkable thing shown in Table 4.12 is that those who are not sure of the reliability of students' test scores all believe that their uncertainty results from the students' prior preparation for test tasks.

Personal details about the respondents from questions 9 and 10 provide important facts:

- Half of the respondents have taught English for over 8 years, 5 – over 5 years, and only one – over 11 years.
- Only one of the subjects has attended a course or a workshop on language testing in Vietnam.

**4.2.2 Analysis of the Results**

This sub-section analyses the data gathered from the questionnaire survey. The information revealed in Tables 4.5 to 4.12 above can be generalized as follows: (1) most of the teachers are unable to distinguish the two functions of spoken language, (2) the majority of the teachers do not have a sufficient understanding of communicative language testing, (3) quite a number of the teachers fail to recognize the difficulty level of test tasks graded according to communicative stress, (4) the teachers have an inadequate understanding of oral language testing, and (5) the speaking tests in use lack authenticity.

Firstly, most of the teachers fail to distinguish the two different functions of spoken language in their own assessment criteria as discussed in 2.1 (Chapter 2). In particular, for Questions 1 & 2 (Tables 4.5 & 4.6), 11 out of 12 subjects give the lowest priority to option (g), that is the ability to speak at length. When taking transactional long turns, students are required to show their ability to express and convey their ideas at length. Nevertheless, the teachers at TNU hardly expect to see this feature in their students' oral test performance.

Secondly, the notion of communicative language testing is not fully grasped and soundly engraved on many of the teachers' mind. In this preferably applied approach to language testing as mentioned in 2.2 (Chapter 2), accuracy of language forms has turned out to be given a lower priority than communicative effectiveness in assessment. As revealed in Tables 4.5 and 4.6, the majority of the teachers, in their assessment criteria, give a high priority to options (a), (b) and (c) of both Questions 1 & 2. In particular, for Question 1, option (a) is given such a high priority by 4 subjects, (b) by 5 and (c) by 7; and for Question 2, option (a) by 5, (b) by 8 and (c) by 8. This implies that on average half of TNU English teaching staff give linguistic accuracy a high priority when evaluating their students' production of spoken language.

Thirdly, as shown in Table 4.7, for Question 3 option (a) has been selected by only 4 subjects, which means that the majority are able to realize the need to include more than one task or elicitation technique in a test of overall speaking ability as discussed in 2.3.2 (Chapter 2). However, in their selection of proper elicitation techniques or tasks for each level, they reveal their inability to recognize the difficulty level of the given elicitation techniques according to communicative stress (Figure 2.1 – Chapter 2), particularly the difficulty of the tasks themselves. To prove this, as shown in Table 4.8, only 2 out of 12 subjects have chosen (d) (the correct answer) for the first level and (c) (the correct answer) for the second level while 6 subjects have chosen (g) (the wrong answer) for the first level.

Besides, this failure is more strongly confirmed by the information gathered from Question 5. As shown in Table 4.9, only 2 out of 12 subjects have chosen (e) (the correct answer) for the first level, 3 have chosen (d) (the correct answer) for the second level, and 2 have chosen (h) (the correct answer) for the third level whereas up to 9 of them have chosen (c) (the wrong answer) for the second level, and 10 have chosen (d) (the wrong answer) for the third level. This result suggests that most of TNU English teaching staff are unable to recognize the difficulty level based on the relationship within each specified task as mentioned in 2.1 (Chapter 2).

Fourthly, as regards all the steps to be sufficiently considered in the development process of speaking tests as described in 2.4 (Chapter 2), only half of the subjects (6/12) have selected option (e) (Table 4.10), and the other options have not been chosen by all the subjects. This partly reveals the subjects' incomplete grasp of language test development. As shown in Table 4.7 (Question 3), there is a great agreement on combining more than one kind of elicitation techniques in a test of overall speaking ability (8 subjects/12). However, in relation to the requirements in the development process of a speaking test, half of them have failed to recognize the importance of identifying the number of test tasks or elicitation techniques included in a speaking test.

Lastly, concerning the usefulness of the oral tests in use at TNU, 10 out of 12 subjects, as revealed in Table 4.11 (Question 7), do not entirely believe that their marks given on the students' oral test performance could reveal the students' actual speaking ability. These 10 teachers have all chosen option (a), as shown in Table 4.12, which implies that in most of the oral tests the students were informed of the topics or test questions during the class time, and they were well prepared for what they were going to speak before the real tests actually occurred. Such speaking tests can thus be said to lack authenticity, defined as 'the degree of correspondence of the characteristics of a given language test task to a target language use task' (Bachman & Palmer, 1996, p. 23), because they do not represent real life language use. Therefore, many of the teachers (7/10) think that the elicitation techniques commonly used in most of the tests are unable to elicit the students' actual speaking ability (option c). In addition, half of them (5/10) have selected the other two reasons for their low confidence in their own given marks. For the first reason, perhaps they think the students could even guess and prepare what they could be asked about as in many cases the topics or test questions were given beforehand. For the second reason, they have had no specified assessment criteria or guideline to base their marking on as described in 4.1.1. As a result, they have had to 'design' their own criteria.

To sum up, the analysis of the data collected from the survey shows that TNU English teaching staff have obtained a limited understanding of oral skill assessment. Obviously, their limited and insufficient grasp of oral language testing may probably lead to their low confidence in their scores given on their students' oral test ability. Also, their improper practices of developing speaking tests critically evaluated in 4.1.3 must be an inevitable consequence of their incomplete perceptions of speaking skill assessment.

**4.3 Summary**

Chapter 4 has described the current process of speaking test development at TNU, presented the results collected from the observation of an end-of-term speaking test and from the survey of the staff's perceptions of oral language testing. Then the evaluation of the current oral test development process and the analysis of the survey result help to reveal that TNU staff's superficial knowledge of oral language testing surely results in their inappropriate practices of developing oral language tests. The study has proved to be beneficial as it helps TNU staff to find out their strengths and weaknesses, and they can therefore identify room for improvement.

## CHAPTER 5: RECOMMENDATIONS AND CONCLUSION

In chapter 4, the results of the study have been presented and analytically discussed in order to find out strengths and weaknesses of TNU oral language testing. The discussion indicates two main findings: first, the current practices are far from being consistent with the theoretical framework of test development, and second, TNU teaching staff have gained limited and insufficient knowledge of oral language testing. These findings serve as the basis for following recommendations regarding standardisation of TNU oral testing practices. This chapter (1) makes several practical recommendations for TNU oral testing practices, and (2) provides a conclusion ending the thesis.

**5.1 Recommendations for TNU Oral Testing Practices**

The findings of the study presented above imply that current oral testing at this institution really need to be improved and standardised in order to gradually increase the training quality of the institution as a whole and of the English Section in particular.

TNU staff's lack of sufficient competence in speaking skill assessment is one of the main drawbacks resulting in their improper practices of oral test development. Since it is TNU English teaching staff who are, at this institution, the most proficient in English teaching, and directly involved in both the teaching and the testing of speaking skill, it is essential that they be aware of the need to become more competent in developing speaking tests. This thesis recommends using Bachman & Palmer's framework for test development (See 2.3 – Chapter 2) for the context of oral ability assessment as a theoretical basis for developing speaking tests at TNU.

Based on Bachman & Palmer's theoretical framework for test development discussed in 2.3 – Chapter 2 – and on the weaknesses of TNU current oral testing practices analysed in Chapter 4, this thesis makes seven particular recommendations, 5 of which are meant to be used to improve the whole present development process of speaking tests, and the other two of which are directly involved in the test operationalization, namely a set of TLU tasks for TNU first-year students and two sample achievement tests for first-year students. This section starts with suggestions for improving the test development process as a whole and ends with practical applications to the operationalization of speaking tests for first-years students.

### 5.1.1 Recommendations for TNU Development Process of Achievement Speaking Tests

The Pedagogy Department or the English Section should produce an official document including the following suggestions this thesis attempts to make as the first effort to standardise TNU current oral language testing. These recommendations for improving the current development procedure include:

1. A rating/level scale
2. A blueprint for development of TNU achievement speaking tests
3. A standardisation meeting
4. A supportive test taking environment
5. Use of test results for teaching evaluation

#### 5.1.1.1 Rating/Level Scale

At this institution, as described in 4.1 – Chapter 4, there has never been an official level scale in English Section training programme in general, for speaking skill in particular. Now a specified level scale is the prerequisite for oral test development, therefore. Regarding to the level scale, the training of speaking skill at TNU includes 240 contact hours explicitly distributed to 6 terms, the first 4 of which are each allotted 45 contact hours, and the rest 2 of which have 30 contact hours. At the beginning of the course the students are supposed to be false-beginners, i.e. at post-elementary level in terms of speaking skill, since most of them, from rural areas and small towns, had no chance of exposure to English, and their English learning at high school just focussed on grammar, structure, vocabulary and reading skill. After 240 contact hours of training, the freshmen are expected to reach upper-intermediate level.

This thesis suggests using the level scale introduced in 2.4.1 on page 17 for TNU speaking skill training course as well as its oral language testing. This level scale, as maintained in 4.1.3 (Chapter 4), helps teachers and examiners to identify the level that best fits their students' proficiency level, and thus design adequate test tasks that are valid for their designated purpose. Obviously, use of this level scale helps to increase test validity.

| Elementary | Pre-intermediate | Intermediate | Upper-intermediate |
|---|---|---|---|
| -introduce oneself and others.-ask and answer questions about personal | -communicate in simple and routine tasks requiring a | -use the language in most situations likely to arise when | -interact with a degree of fluency and spontaneity |

| details such as where he/she lives, people he/she knows and things he/she has.-interact in a simple way provided the other person talks slowly and clearly and is prepared to help.-use very simple expressions related to areas of most immediate relevance (e.g. very basic family information, shopping, local geography, employment). | simple and direct exchange of information on familiar and routine matters.-dsecribe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.-simply talk about familiar matters regularly encountered in work, school, leisure, etc. | travelling in an area where the language is spoken.-make a simple connected presentation on topics which are familiar or of personal interest.-describe experiences and events, dreams, hopes and ambitions.-briefly give reasons and explanations for opinions and plans. | that makes regular interaction with native speakers without strain for either party.-make a clear and detailed presentation on a wide range of subjects.-give opinions on topical issues and explain the advantages and disadvantages of various options. |
|---|---|---|---|

**5.1.1.2 Blueprint for Development of Achievement Speaking Tests at TNU**

A blueprint which has been discussed in 2.3.2 (Chapter 2) includes the number of test tasks/parts and specifications of each task. The blueprint is suggested as major guidelines for TNU staff to construct their speaking tests.

As discussed in 2.4.2 (Chapter 2), a test of speaking ability that enables assessors or examiners to elicit a test taker's overall oral proficiency should consist of at least two tasks or elicitation techniques. Undoubtedly, a speaking test making use of two or more tasks or elicitation techniques is said to have construct validity and reliability. Regarding the speaking tests of several popular published exams, a test of overall oral ability always has two or three tasks/parts. For instance, BEC (Business English Certificates) and IELTS (International English Language Tests) have a three-task speaking section, or Let's Talk presents speaking tests involving two tasks/parts. Thus, it would be advisable to design a blueprint for speaking tests at TNU including two tasks or elicitation techniques.

The following are the suggested components of the blueprint for an achievement oral test.

1. Test structure

1.1 Number of tasks/parts:2 tasks

Language, as discussed in 2.1 (Chapter 2), has two functions, so assessment of students' ability to use the language orally is to involve these both functions. Furthermore, length of spoken language production is also the basis for this kind of assessment (See Figure 2.1, page 6).

Task 1:Interactional and transactional short turns

The purpose of this task is to evaluate students' progress in taking interactional and transactional short turns.

Task 2:Transactional long turns

The purpose of this task is to evaluate students' progress in taking transactional long turns.

1.2 Relative importance of the tasks

This importance ranges on the continuum of spoken language production (Figure 2.1) according to students' levels of language proficiency. For instance, in a test for first-year students, Task 1 is more important than Task 2.

2. Test task specifications

The purpose of the task

The specified components of oral ability to be tested

The place where the task occurs

Specified and understandable instructions

Expected duration of task performance

Areas of linguistic, pragmatic and topical knowledge adequate

Marking key

Concerning criteria for scoring, based on the discussion on a marking key in 2.4.3 (Chapter 2), the researcher suggests combining the two models of mark categories, i.e. the traditional model and the model of performance criteria, since doing so means that linguistic accuracy is not neglected and objectives of the course or the teaching are not thus neglected either. That is to say, we as language teachers never want to neglect linguistic forms when instructing students. Therefore, this thesis recommends the marking scales adapted from PET Speaking Test by University of Cambridge Local Examinations Syndicate, representing a combination of these two model. These marking scales are used in the marking key of two sample achievement tests introduced in the next sub-section 5.1.2

### 5.1.1.3 Standardisation Meeting

As discussed in 4.1 – Chapter 4, before TNU oral tests are actually administered, no discussion taking place to reach agreement on how to mark each question/task among the group of assessors is proved to also affect test reliability. It would thus be strongly advisable that the English Section should potentially hold just a short standardisation meeting in order to ensure that enough discussion will take place for all examiners to understand thoroughly the level scale and the procedures for scoring. Of course, at this meeting the level scale (discussed in 5.2.3) and the marking key for each test task (discussed in 5.2.5) are really needed. As such a meeting is possible, it will help to increase test reliability.

### 5.1.1.4 Supportive Test Taking Environment

As mentioned in 4.1 – Chapter 4, most of the speaking tests have taken place in noisy rooms, which surely affects students' test performance and thus reduces reliability of the tests. Therefore, in order to ensure that a test is reliable, it is crucial to maintain a supportive environment throughout the test. In particular, examiners and administrators should avoid distractions due to temperature, noise, excessive movement, and so on, and provide a comfortable room for those students waiting for their turn. In my opinion, it would be feasible for TNU to maintain such a testing environment.

### 5.1.1.5 Use of Test Results for Teaching Evaluation

At TNU, as previously described and analysed, students' test scores have never been used to either determine the effectiveness of instructional programs or make any improvement in teachers' teaching methods and materials, which reveals that TNU staff do not exploit test potential and usefulness to improve their teaching as well as their testing. This sub-section hence aims at helping the concerned staff to develop a plan for teaching evaluation based on test scores collected.

In particular, teachers should first keep a list of scores in order to evaluate students' achievement in general, i.e. to find out whether the instruction has helped students develop this skill. Then they should specify typical problems impeding the majority of students' performance during test administration in order to find out suitable strategies to promote effective learning, i.e. to modify teaching methods and materials.

With the test scores, as Madsen (1983, p. 5) maintains, teachers might well ask themselves whether their teaching is effective, which is particularly suggested using the following questions:

'1. Are my lessons on the right levels? Or am I aiming my instruction too low or too high?

2. Am I teaching some skills effectively but others less effectively?

3. What areas do we need more work on? Which points need reviewing?

4. Should I spend more (or less) time on this material with next year's students?'

And test administration can 'provide insights into ways that we can improve the evaluation process itself.'

'1.Were the test instructions clear?

2. Did the test cause unnecessary anxiety or resentment?

3. Did the test results reflect accurately how my students have been responding in class and in their assigned work?'

**5.1.2 Practical Applications to the Operationalization Process of Speaking Tests for First-Year Students**

The following two recommendations hopefully help the test writers to visualize how to write a speaking test for TNU language students in general and for first-year students in particular.

**5.1.2.1 Suggested Tasks in the TLU Domain for Inclusion in Speaking Tests for First-Year Students**

One of the important steps in test development, as described in 2.3.1 (Chapter 2), is identification of tasks in the TLU domain for selection of actual test tasks included in a test. TLU tasks must be selected based on the teaching content, yet at TNU each teacher in charge of speaking skill training has used their own teaching materials. No speaking skill syllabus or teaching material has been officially approved; a syllabus for this skill is now under development and revision. Therefore, TLU tasks can not be typically chosen from all the materials unofficially in use, and all the following suggested TLU tasks, adapted from a course book named 'English for International Communication' by Richards (2002), are intended to partially help the author to design two sample speaking tests for first-year students or for elementary and post-elementary levels in the level scale suggested in 5.1.1.1. TLU tasks are suggested as follows:

- Introducing oneself or someone
- Exchanging personal information
- Describing school and house
- Talking about families and family members
- Describing family life
- Talking about daily activities
- Talking about likes and dislikes
- Buying and selling things, Talking about prices, ordering a meal
- Asking about and describing locations of places
- Asking about and describing people's appearance
- Asking about and describing objects
- Making invitations and excuses, accepting and refusing invitations
- Talking abilities
- Talking about past experiences and events
- Making comparisons
- Asking for and giving advice
- Asking for and giving suggestions
- Taking and leaving messages

- Describing changes
- Talking about plans for the future

**5.1.2.2 Two Sample Achievement Speaking Tests for First-Year Students**

The two achievement speaking tests designed below are based on the blueprint suggested in 5.1.1.2

Sample achievement speaking test for first-year students – Term 1

Task 1: Conversation between the assessor and each student about personal information.

Purpose of the task:

to assess the students' ability to interact in typical daily situations

Specified components of speaking ability to be tested:

The ability to talk about themselves and use social language in common interactions

The place where the task occurs:

In the classroom

Expected duration of task performance:

About 2 minutes

Specific and understandable instructions:

Each student starts the dialogue with the assessors. The assessors ask each student questions about himself/herself and about his/her family.

Areas of linguistic, pragmatic and topical knowledge adequate

Simple and common vocabulary and grammar

Simple functions such as greeting, agreeing or disagreeing, and easy description.

About oneself and his/her family

Marking key

Students are assessed on their own performance according to the criteria in the marking scales in the table 5.1

MARKING SCALES – TASK 1 (6 marks out of 10)

| Marks | Fluency | Accuracy and Appropriacy of Language | Pronunciation | Task Achievement |
|---|---|---|---|---|
| 6 | Occasional hesitations, but not such as to impede communication | Meaning is conveyed despite noticeable structural inaccuracies, lack of vocabulary | Generally easy to understand despite L1 accent | Tasks dealt with adequately |

| | | | | |
|---|---|---|---|---|
| 5 | Hesitation often demands unreasonable patience of listener. | Meaning occasionally obscured by structural inaccuracies and limited vocabulary | L1 interference occasionally causes difficulty in understanding | Limited ability to deal with tasks |
| 4-3 | Speech very disconnected and difficult to follow | Frequently incomprehensible because of limited vocabulary and numerous structural errors | Frequently impossible to understand | Ineffective handling of tasks |
| 2-1 | No connected speech. | Incomprehensible because of insufficient vocabulary and gross structural errors | Impossible to understand | Unable to deal with tasks |

Table 5.1: The Marking Scales for Task 1 of the Sample Term 1 Achievement Speaking Test

Task 2: Describe your normal day

Purpose of the task:

to assess the students' ability to use English to take a bit transactional long turns, ie. communicate some information.

Specified components of speaking ability to be tested:

The ability to make a description through a short oral presentation.

The place where the task occurs:

In the classroom

Expected duration of task performance:

About 3 minutes

Specific and understandable instructions:

The student tells the assessors about the main activities you normally do during the day. Your talk is about 100 words or less.

Areas of linguistic, pragmatic and topical knowledge adequate

Simple and common vocabulary and grammatical structures

Functions such as starting and closing a presentation

About oneself and common daily activities

Marking key

Students are assessed on their own performance according to the criteria in the marking scales in the table 5.2 on page 68

MARKING SCALES – TASK 2 (4 marks out of 10)

| Marks | Fluency | Accuracy and Appropriacy of Language | Pronunciation | Task Achievement |
|---|---|---|---|---|
| 4 | Occasional hesitations, but not such as to impede communication | Meaning is conveyed despite noticeable structural inaccuracies, lack of vocabulary | Generally easy to understand despite L1 accent | Tasks dealt with adequately |
| 3 | Hesitation often demands unreasonable patience of listener. | Meaning occasionally obscured by structural inaccuracies and limited vocabulary | L1 interference occasionally causes difficulty in understanding | Limited ability to deal with tasks |
| 2 | Speech very disconnected and difficult to follow | Frequently incomprehensible because of limited vocabulary and numerous structural errors | Frequently impossible to understand | Ineffective handling of tasks |
| 1 | No connected speech. | Incomprehensible because of insufficient vocabulary and gross structural errors | Impossible to understand | Unable to deal with tasks |

Table 5.2: The Marking Scales for Task 2 of Sample Term 1 Achievement Speaking Test

Sample achievement speaking test for first-year students – Term 2

Task 1: The student starts the conversation to make the acquaintance of the assessor(s) at a party and then makes an invitation or an offer.

Purpose of the task:

to assess the students' ability to interact in usual daily situations

Specified components of speaking ability to be tested:

The ability to introduce themselves and to use social language in common interactions

The place where the task occurs:

At a party

Expected duration of task performance:

About 2 minutes

Specific and understandable instructions:

Start the dialogue and make the acquaintance of the assessor(s). Then invite the assessor(s) to have something or to do something.

Areas of linguistic, pragmatic and topical knowledge adequate

Simple and common vocabulary and grammar

Functions such as greeting, addressing, making someone's acquaintance and making invitations.

About oneself and normal daily meetings

Marking key

Students are assessed on their own performance according to the criteria in the marking scales in the table 5.3 below

MARKING SCALES – TASK 1 (5 marks out of 10)

| Marks | Fluency | Accuracy and Appropriacy of Language | Pronunciation | Task Achievement |
|---|---|---|---|---|
| 5 | Occasional hesitations, but not such as to impede communication | Meaning is conveyed despite noticeable structural inaccuracies, lack of vocabulary | Generally easy to understand despite L1 accent | Tasks dealt with adequately |
| 4 | Hesitation often demands unreasonable patience of listener. | Meaning occasionally obscured by structural inaccuracies and limited vocabulary | L1 interference occasionally causes difficulty in understanding | Limited ability to deal with tasks |
| 3 | Speech very disconnected and difficult to follow | Frequently incomprehensible because of limited vocabulary and numerous structural errors | Frequently impossible to understand | Ineffective handling of tasks |
| 2-1 | No connected speech. | Incomprehensible because of insufficient vocabulary and gross structural errors | Impossible to understand | Unable to deal with tasks |

Table 5.3: The Marking Scales for Task 1 of Sample Term 2 Achievement Speaking Test

Task 2: Talking about your next summer vacation

Purpose of the task:

to assess the students' ability to use English to take a bit transactional long turns, ie. to communicate some information.

Specified components of speaking ability to be tested:

The ability to make an oral presentation on their future intentions

The place where the task occurs:

In the classroom

Expected duration of task performance:

About 4 minutes

Specific and understandable instructions:

Tell the assessors about your next summer vacation such as where to go, what to do, why you want to do so and how long to stay there. Your talk is about 150 words or less.

Areas of linguistic, pragmatic and topical knowledge adequate

Simple and common vocabulary and grammatical structures

Functions such as starting and closing a presentation

About oneself, future plans and hobbies

Marking key

Students are assessed on their own performance according to the criteria in the marking scales in the table 5.4 on page 71

MARKING SCALES – TASK 2 (5 marks out of 10)

| Marks | Fluency | Accuracy and Appropriacy of Language | Pronunciation | Task Achievement |
|---|---|---|---|---|
| 5 | Occasional hesitations, but not such as to impede communication | Meaning is conveyed despite noticeable structural inaccuracies, lack of vocabulary | Generally easy to understand despite L1 accent | Tasks dealt with adequately |
| 4 | Hesitation often demands unreasonable patience of listener. | Meaning occasionally obscured by structural inaccuracies and limited vocabulary | L1 interference occasionally causes difficulty in understanding | Limited ability to deal with tasks |
| 3 | Speech very disconnected and difficult to follow | Frequently incomprehensible because of limited vocabulary and numerous structural errors | Frequently impossible to understand | Ineffective handling of tasks |
| 2-1 | No connected speech. | Incomprehensible because of insufficient vocabulary and gross structural errors | Impossible to understand | Unable to deal with tasks |

Table 5.4: The Marking Scales for Task 2 of Sample Term 2 Achievement Speaking Test

**5.2 Conclusion**

Current oral language testing practices at TNU have been claimed to be very problematic, so this study is intentionally carried out to investigate the present practices. In reality, oral testing at this institution is far from being consistent with the theory in language testing. This finding is based on the result of the study which is aimed at evaluating the practices and the staff's perceptions of oral testing.

The detailed review of the existing practices provides the basis for the analytical evaluation in order to identify the strengths and weaknesses. Also, a questionnaire survey helps to elicit the staff's understanding of oral testing – the cause of the current practices.

The analysis of the study results shows that the present problem stems from the staff's insufficient knowledge of oral testing. All the tests of speaking ability were inappropriately constructed and inadequately administered.

As a result, the findings of the study help to justify the claim and several practical recommendations are proposed in terms of the procedures and activities involved in oral test development.

Obviously, the study as a whole can be considered significant as it provides two main practical contributions towards TNU context of language testing as follows.

Firstly, the study has applied the theory of language testing, a science, to TNU language teaching with the purpose of ensuring and increasing the training effectiveness, namely evaluating and promoting professionalism in language training at TNU. In particular, the theory reviewed in the study surely provides a reasonable foundation on which TNU staff's testing of oral proficiency can be based; thus, it is hoped that the review will assist the staff in better understanding the testing field.

Secondly, the study has made seven practical recommendations for TNU staff's development process of speaking tests. Five of the recommendations are concerned with relevant applications in relation to the theoretical considerations in test development process as a whole, which is aimed at providing the concerned staff the guidelines for developing oral tests. The other two recommendations as a case study are related to the operationalization of a particular speaking test.

The findings of the study also help to make suggestions for further research. The fact is that no specified syllabus is designed for speaking ability teaching at TNU and the same circumstance for the other three subjects – listening, reading and writing. Therefore, testing of the other abilities is really problematic, and this kind of research is necessary to increase the quality of the training program of the English Section as well as of the institution as a whole. Areas of research should be concerned with development of tests of the other abilities or skills, and development of syllabuses for teaching of all the four skills or abilities.

In conclusion, the study will hopefully be taken in account during the process of oral test development of TNU staff – test designers or teachers, assessors or examiners as well as test administrators. However, one limitation of this thesis is that the researcher has been unable to find out the method of quantifying the results gathered from the survey of the staff's perceptions of oral language testing.

## REFERENCES

Alderson, J. C., Clapham, C. & Wall, C. (1995). Language test construction and evaluation. Cambridge: Cambridge University Press.

Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L. F & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford: Oxford University Press.

BEC – Business English Certificates Handbook (Revised syllabus). University of Cambridge Local Examinations Syndicate.

Brown, H. D. (1994). Principles of language learning and teaching. Englewood Cliffs, NJ: Prentice Hall Regents.

Brown, H. D. & Gonzo, S. (1995). Readings on second language acquisition. Englewood Cliffs, NJ: Prentice Hall.

Brown, G. & Yule, G. (1983). Teaching the spoken language. Cambridge: Cambridge University Press.

Butler, F. A. et al (2000). TOEFL 2000 speaking framework: A working paper. (TOEFL Monograph Series Report No. 20). Princeton, NJ: Educational Testing Service.

Catt, C. (2003). IELTS Speaking: Preparation and practice. London: Longman.

Celce-Murcia, M. & Olshtain, E. (2000). Discourse and context in language teaching. Cambridge: Cambridge University Press.

Cohen, A. D. (1996). Assessing language ability in the classroom. Massachusetts: Heinle & Heinle Publishers.

Davies, P. (2000). Success in English teaching. Oxford: Oxford University Press.

Gillham, B. (2000). Developing a questionnaire. London and New York: Continuum.

Heaton, J. B. (1988). Writing English language tests. New York: Longman.

Heaton, J. B. (1990). Classroom testing. New York: Longman.

Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.

Madsen, H. S. (1983). Techniques in testing. Oxford: Oxford University Press.

Metre, D. V. (2003). Let's talk – Testing packet. Cambridge: Cambridge University Press.

McNamara, T. (2000). Language testing. Oxford: Oxford University Press.

Nachmias, C. F. & Nachmias, D. (1996). Research methods in social sciences. New York: Oxford University Press.

Nunan, D. (1991). Language teaching methodology: A textbook for teachers. Englewood Cliffs, NJ: Prentice Hall.

O'Malley, J. M. & Pierce, L. V. (1996). Authentic assessment for English language teachers. Addison-Wesley.

PET Speaking Tests by University of Cambridge Local Examinations Syndicate.

Richards, J. C. (2002). English for international communication. Cambridge: Cambridge University Press.

Underhill, N. (1987). Testing spoken language: A handbook of oral testing techniques. Cambridge: Cambridge University Press.

Ur, P. (1996). A course in language teaching. Cambridge: Cambridge University Press.

Vu Thi Phuong Anh & Nguyen Thi Kim Thu (2003). He thong dinh chuan trinh do ngoai ngu cua Hoi dong Chau Au. In Hochiminh City University of Social Sciences and Humanities Workshop Record: Khoa hoc Xa hoi va Nhan van trong boi canh hoi nhap Quoc te (p. 270-277). Hochiminh City: Hochiminh City University of Social Sciences and Humanities.

## APPENDIces

**Appendix 1: Three Achievement Speaking Tests Used at TNU**

TEST 1: ENGLISH SPEAKING TEST

Class: English K2000 – Term 1

Students are supposed to give a presentation about one of the following topics, then answer examiners' two or three additional questions.

1. Please introduce yourself. We would like to know about your birthplace, your age, and your family. How does your family help you in your study?
2. Could you tell us about your daily activities and then your freetime activities? What are your likes and dislikes?
3. Could you tell us about your new class and your new school? How did you feel when you first arrive here?
4. Would you please describe the appearance of a special friend in your class? Why is he/she special to you?
5. Who is the monitor of your class? What's he like? What do you think of him as a monitor?
6. Please describe your mother. Do you look like her? What kind of clothes does she often wear? Do you have the same way of dressing?
7. Please describe your father's character. What is his job? Do you often talk with him about your study and your friends?
8. Where are you from? Could you tell us about some of the interesting places in your hometown?
9. Please talk about your hometown. We would like to know about its location, the weather and the environment there.
10. Could you tell us about life of people in your hometown? (population, their main jobs, their standard of living, their character...)
11. What do you remember most about your childhood? Did you have a happy child hood?
12. You are now a first-year student. Is there any difference between your life now and life when you were a highschool pupil?
13. When do you think Christmas season begins? Could you describe the atmosphere around when Christmas is coming?
14. Who is the most famous person at Christmas? Can you describe him? What kind of gift would you like him to give you?
15. What did you do last Christmas? (Before, on Christmas Eve, on Christmas Day.) How did you feel?
16. Lunar New Year Holiday is coming. How do you feel? What do you like most about it?
17. What do you often do before the Lunar New Year Holiday? How do you feel at the very moment when the old year is out and the new year is in? what do you often do then?
18. Could you tell us about some of the Vietnamese customs concerning the Tet Holiday?
19. Are you superstitious? Could you give us some examples of your own superstition or your family superstition or the superstition of the Vietnamese people during the Tet Holiday?
20. What do you often do on the three days of Tet? How do you feel on these days?

TEST 2: ENGLISH SPEAKING TEST

Class: English K2000 – Term 2

1. Do you often cook the meals for your family? Why? Why not? Do you think that knowledge of cooking is essential? Why so? Why not?
2. Some people say that cooking is the duty of women and girls only, men and boys don't need to learn how to cook. Do you agree with them? Why? Why not?
3. 'I am the only child of my parents so though I'm a girl, I don't have to do any housework. I just have to study. In my freetime, I can go out with my friends, listen to music or watch TV. Everything in the house is done by my mother because she doesn't go to work,' said Mai, a 19-year-old student.

What do you think about Mai's behaviour at home and her duty to her family?

1. Some people say that cooking is a waste of time because fast food is available now, and that we should save time to do other necessary things. What do you think about this idea? Completely agree? Partially agree or disagree? Why?
2. Read the dialogue below and answer the questions given.

A: I've been looking at your brooch. It's very unusual. Where did you get it?

B: I got it in Malaysia.

A: Oh, did you? How long were you there? By the way, I'm John Gooch,...

B: I'm Sylvia Martin. I was there for three years actually.

A: Really? That must have been a wonderful experience. Wht did you enjoy most?

Questions:

1. Do you think the two people in the conversation have known each other before? How do you know that?

2. How does John Gooch start the conversation? What is the common way to start a conversation?

   1. Many students have meals in the school canteen or in a café near their school or where they are living, but there are some students who cook for themselves. Which do you think is better?
   2. Read the dialogue below and answer the questions given.

A: May I introduce myself? I'm Robert Munns.

B: How do you do? I'm Tina Morley.

A: How do you do?

Questions:

1. What is the relation between Robert Munns and Tina Morley? How do you know that?

2. How do Robert and Tina introduce themselves to each other? What other information should people give when introducing each other?

   1. In the exam room, what do you do if you do not know or are not sure of the answer? If your friends have different choice or answer to a question from yours, will you change your mind?
   2. As a student of English, what English subjects or skills are you learning? Which do you find the most difficult? Do you know the reason why? What do you think you have to do to learn that subject or skill better?
   3. Read the dialogue below and answer the questions given.

A: Mr. Granger, I'd like you to meet Nick Thomas, from our Boston office.

B: How do you do, Nick.

C: Please to meet you, Mr. Granger.

B: Please call me Philip.

Questions:

1. What information should we mention when introducing someone to another?

2. what is the correct order when introducingsomeone to another? (In terms of age, sex, and position in society)

   1. What is the role of grammar in learning English? What is its relation to the other skills?
   2. Can you explain why a student cannot express himself or herself or be understood by others although his/her knowledge of grammar is good? He/She can do grammar exercises well and his/her writing skill is rather good.
   3. Are you worried or nervous when examinations are coming? What are you worried about? In order not to be worried about taking exams, what do you think you have to do during the semester?
   4. Should you help your friends in the exam room? Why? Why not? How should you help your friend(s)?
   5. What time do you go to bed every day? Do you go to bed later than usaul when examinations are coming? Some students stay up too late or even the whole night before the exam. Do you think that is a good way of learning? Why? Why not?

6. Some students often feel sleepy when they are learning hard for the exams. How about you? What do you do if you feel sleepy when learning, going to bed or trying to do something to be awake? Why do you do so?
7. Do you think that examinations are important in the process of teaching and learning? Why? Why not?
8. Have you ever failed in your exams? If yes, how did you feel and what did you do after that to improve the situation? If not, how do you think you will feel? What will you do after that?
9. If your friends or younger brothers or sisters ask you to give them some advice to prepare well for examinations, what is your advice?
10. Talk about your summer holiday this year. What are you going to do after taking the exams?

TEST 3: ENGLISH SPEAKING TEST

Class: English K2000 – Term 3

1. What job would you like to have after you graduate from university? State the reasons why you like that job.
2. Between the job of a teacher of English and a tourist guide, which one would you prefer? State the reason for your choice.
3. Do you want to work as a secretary in an office? Why? Why not?
4. Between the job of a secretary in a big company and an air-hostess, which one would you prefer? State the reasons for your choice.
5. What kindds of job do you think a student of English can have? State what they will use English for.
6. Many people say reading is the best way of widening our knowledge; do you agree with them? Why? Why not?
7. What do you think a student of English should read? Let us know your habit of reading (what you read? When? How often?, what you think, ect.)
8. Do you agree that the pleasures of reading are varied according to age, personality, jobs, ect. Give examples to illustrate your idea.
9. Have you ever found some money in a library book? If yes, what did you do with that sum of money? If no, what would you do if you found 100,000 VND in a library book?
10. What would you do and how would you feel if you were invited to the rector's party?
11. If you could make three wishes, what would they be?
12. Read the dialogue below and state if you agree with Peter's idea: 'digging garden is not a jod for a University graduate,'

Peter: Why don't you get a decent job for a change?

Dick: But I like my job.

Peter: Look, digging garden is not a job for a University graduate.

Dick:But the money's not bad and there's plenty of fresh air.

Peter:If I were you, I'd go on some kind of course – teaching, accountancy.

Dick:Accountancy? Anything but that. It's too boring.

Peter:Come on, you really must think of the future. Why don't you just write a few application froms?

Dick:I'll tell you what. I'd really like to be a doctor.

Peter:Well, you should think very seriously about that. It means a lot of study, and then working all sorts of hours.

Dick:Yes, may be. But the idea appeals to me.

Peter:Well then, you ought to get more information about it as soon as possible.

1. Read the dialogue below and then paraphrase it in your own words, mentioning the relationship between them, where they are and what they are talking about, and so on.

Paola:You must take some rest. You've been working too much hard.

Mary:But how can I? The deadline is Friday.

Paola:Come on, couldn't you take the afternoon off?

Mary: Well, if you really think so.

Paola:I really think you should. We can manage without you.

1. Your friend tells you he/she met a wonderful woman/man yesterday and is getting married next week. What do you think and what will you say about his/her sudden decision?
2. You have got tickets for a film. At the last moment your girlfriend/boyfriend rings up and says she/he has a headache and can't come. What will you do and say to him/her?
3. In your daily life, from whom do you often have to ask for permission? What permission you may want to ask for from your teachers? If you have a dental appointment and you need tomorrow off, what do you say to your teacher?
4. In what situations do people often make comlaints? Have you ever made a complaint? What was it about?
5. What sentences can you say when you want to ask someone for a lift to the station?
6. Read the dialogue below and state if you agree with Peter's idea about the work of a doctor.

Peter: Why don't you get a decent job for a change?

Dick: But I like my job.

Peter: Look, digging garden is not a job for a University graduate.

Dick:But the money's not bad and there's plenty of fresh air.

Peter:If I were you, I'd go on some kind of course – teaching, accountancy.

Dick:Accountancy? Anything but that. It's too boring.

Peter:Come on, you really must think of the future. Why don't you just write a few application froms?

Dick:I'll tell you what. I'd really like to be a doctor.

Peter:Well, you should think very seriously about that. It means a lot of study, and then working all sorts of hours.

Dick:Yes, may be. But the idea appeals to me.

Peter:Well then, you ought to get more information about it as soon as possible.

1. Have you ever got a letter from a person you don't know? If yes, what did you do with it? If no, imagine what you would do if you got a love letter from a person you didn't know?


**Appendix 2:**


**Achievement Speaking Test for the Second-Year Students**


**(Term 2 – School Year 2002-2003)**

ĐỀ THI NÓI ANH VĂN 4 - LỚP ANH VĂN K2001

Thí sinh bốc thăm và thuyết trình về 1 trong các đề tài sau: (trong vòng 5 phút)

1. What do you think of the importance of money? Is it always wonderful to have a lot of money?
2. Your opinions about the size of a family? How many children would you like to have? Why?
3. Are there any advantages/disadvantages of living in a multi- generation family? Justify your ideas.
4. If you were born again would you like to have your sex changed? Why/ Why not?
5. What do you think of the teaching career? Why did you choose teaching as your future career?
6. What is your favourite subject? Why do you like it?
7. In your opinion, is examination necessary or should it be given up? Why?
8. In your opinion, what are the roles of a woman in the modern society?


**Appendix 3:The Tapescript of the Test Recorded**

Oral Test Performance of 10 Second-Year students

Student 1:

Good morning everyone. Today I'm in front of you to tell you something about roles of a woman. People said woman are...is heart of the world, so there's no woman the world will not exist. In my opinion, woman is always play an important role in our life, especially in modern life. First, woman help and teach the children. Although the scientists can make a child from test-tubes, but I think the role of a woman can be replaced. Woman... A child is bad or good depend much on the way the mother teach you. If you are a bad mother, your children is not good. And if you are a good mother, your children may be good at other people. .... He will do... He will do his ... well. And the second, woman is always good at doing the housework. It's difficult for men to do all things in his house. But I think it's very easy for a woman to cook some meals, to sweep her house and to take ... of people in her family. ... She not only know how to do it but also know how to do it well. She makes her house more ... and neatly. She takes care of whole her family with good meals. And maybe there is a woman in the house, the house.... the family may be better. Third, in our society I think woman has the same position as men. She can do anything her husband can do. For example, go to university and her profession or and become independent. She shares money to improve her... her family life. She helps her husband with his work. She share all people in work and in .... with her husband. She ... she.... People often say that behind a successful man is a good woman. And I think it's difficult for men to successful without having the help of his wife. These are the reasons why I said that women are always play an important role in our life, especially in modern life. Thank you for your listening.

Student 2:

Today I would like to present my topic – the size of a family. If anyone ask me how many children I would like to have when I ... I will say just one or two children. Why I say so, for I see that small... small family has many advantage than disadvantage. Firstly, having many children I can support ... support all needs of my children. Although in our society to earn much money isn't easy. If I have... having many children I must buy food, clothes ... for them, epecially when they grow up, I must send them to school. School fee for the learning... the learning is... isn't easy. Why I have... if I have one or two children, it is not problem. Secondly, if you are parents in the future you should understand their... the feelings of your children. If you have many children you must do hard to earn money. You have no time to... have no time to share the feeling. It is difficult for you to understand... understand them and... and you don't know the way to teach them become... become good preson. Finally, I think... I think in family have many children it is also having many... having many noise. And children are always playing together. For example, your family has just one TV... Girls want to watch music programme while boys want to watch football programme. They may quarrel together to watch. So... although advantage I think ... I want to... I just want to have one or two children when I married. That's all.

Student 3:

Now I want to tell you about my thinking of money. Nothing is more powerful than money, so money plays an important part in our life. First, if we have money we can buy everything such as we can buy clothes to make more... to make us more beautiful, more confident when going out or standing before a crowd. We can buy

nutritious food to improve... to improve our life. When we are ill we can buy medicine, go to hospital. Secondly, with money we can improve our spiral activities easily. For example, we can buy television... to relax a hard day. With money we can travel... in our... in our summer holiday. Further, we use money not only to meet the basic needs but also... but also to pay... to pay our... investment for education which helps us to... how... to know culturally better. Nowadays many schools, many hospitals being built, but thousands of people are not able to bo to school and... and... and quiting... because their parents have no money. So if we have money we can go to any school, any university we like or we may be sent to abroad to study to further our knowledge. If we... if we are rich we can help the poor and the old people without... without children. However, lack of money our life have many... many difficulties. In fact thousands of people are being dying every day because they have no money to pay their... or they can't buy nutritious food to improve their health. To sum up, money is... indispensible in our life. Thanks to money, it maybe gives us a comfortable life and a cheerful heart. That's all.

The assessor:If you have a lot of money what will you do?

Student 3:If I have a lot of money there are many things I want to do. But the first thing I do... I buy book, I buy English book for me... to... to study better.

The assessor:OK. Thank you.

Student 4:

Hello everybody. My topic is roles of women in the modern society. Women in the modern society play an important role. They not only have to be good at housework but also have to complete the work outside home. It's said that men build the house women make the home. Women have to make their family happy. They always have to do housework completely. The house is always clean when husband and children come home... They have a good meal together. Women also have to take care of her husband and children. After a hard day... after a hard day taking care of wife will take the time with the husband. Women have to teach the children to be good children. Everything will be difficult when we in charge of women's hands. In the old time, women just did only housework and... were not to charge such work, but now they also have to .... with what in society. They have to earn money... Women and men have to take care of their family together. Women can work than men and can get higher position in a company... offices. Being a women we have to ... being a women in the modern society I have to ... complete not only at home but also in society. It's said that women is the heart of the world, so some day there's no woman in the world, I think, the heart of the sky will be .... the heart of the world will be ruined. That's all.

Student 5:

Good morning teachers. My topic is the subject I like best. As you know, up to now I learn.. have learned many subjects, for example mathematics, history. The subject I like best is English. Yes. There are many, I think, there are many reasons to make me to like English, but there are 3 main reasons. The first, with English I can be ability to learn it. That means I can learn it well. Yes. It is very easy for me to understand my teachers said, my teachers teach. And I can... I feel confident when I do exercise... So second, I think, in my opinion, I think with English subject... with English I can read many books... I read and understand many books... example Sunflower... Sunflower magazine. And with English I can talk and understand... tell and understand with foreigners. Maybe I can help them when they lose way or they want to buy something in the market, but they don't know Vietmanese. The third, I... I learn... if I learn English well, I can choose many careers. Yes. As you know, nowadays English in many careers... example in school I can become a teacher, English teacher. In companies I can become a translator, translator, so maybe... To sum up, I choose English... I choose English is my best favourite subject because... I... I feel comfortable when I learn it and with English I can choose many careers. Thank you.

Student 6:

Good morning everybody. Today I'd like to present my topic. The topic is my opinion about the size of family... and... how many children I'd like when I get married. First, my opinion about the size of family. I think small family is always better than big family because when living in small family parents, husband and wife, cam support the needs of family easily. When... only when living in small family we can send our children to good school where... which have the best education and the best training. And... when I married I... two children is enough for me because you and I are parents to be and we have to responsible for the children ... for our children.

For example, we have to taking care of our children's health, and we have to send our children good school. First, we have... we want to send our children to good schools. When we live in small family we have a lot... we go out to earn money and we save a lot of money to send our children to good school. And when we were sent... we are sent to good school we have money to support them to buy some good books, some equipment for learning ... for studying... And about the health, when we live in small family, I think, we have a good taking care of the health, so I'd like to have only two children, one boy or one girl is better, but both of them are the boys or girls is no problem for me. That's all.

Student 7:

Good morning everybody. How do you feel now? I'm a little nervous, but be confident. Yeah, my topic is opinion about the multi family.

The assessor:Multi-generation family.

Student 7:Yes. Living in a multi-generation family, there is plenty of fun, but there are some problems. Nowadays parents have to work all day, so they don't have much time to care for their children. In multi-generation families, grandparents will have done to care for their children... Children can share their troubles, they can share everything that happen in their daily life to their grandparents. And when they meet problems they will have... they will get valuable advice from their grandparents. Grandparents will tell stories, will... they will sing many folksongs and... they will care for them all day when their parents are in work. I think there is plenty of fun, but there are some problems. We can see that a multi-generation family is the one that consists... that consists of three or more generations. And members in such kind of family are not in the same generation, so there are many differences in their opinion and their lifestyle. When old people prefer a quiet life, the young like to live an active and noisy life, together for pop music that too loud for old people. And their grandparents will complain and that makes them not pleased. Nowadays young people want to dress sexually when their grandparents don't want their grandparents... their grandparents to do so. So i'm sorry... when their grandparents don't want their granddaughter or grandson to do so. So they will complain and that makes them not pleased. It's not eassy to compromise their lifestyle and their opinion. So to sum up, I think that everyone has two sides and a multi-generation family is the same. It has plenty of fun, but it also has problems. I think that is... that's all. Thank you for your listening.

Student 8:

Hello everyone. Today... my topic today is teaching career. In society there are many careers, but teaching is the job I like best. I'm going to tell you the reason why I like this job. First, teaching career get high respect from society. It's... it's not only gets respect from people but also gets respect from their... their parents or people around... around them. Second, teaching career... teaching career is the career not ... it less competition and... you just go to class and teach... teach the pupils. You... you must have to worry about whether you get higher position than others. This makes you always feel happy. Besides it, I think, teaching career is the career you... which you... have to learn all life time. It seems to be not good, but all of us have the need to learn. Nowadays the world have many chances, but we always have to learn to improve our knowledge, to teach students. To sum up, teaching career is not only get high position but also it takes less competition than other careers.

Student 9:

Good morning teachers and everybody. Today I would like to tell you the importance of money. We can't live without money. We need money to satisfy... our life. We need money to ... save for expenses, for example in family, in business and other problems. So, in some extent, money play important and necessary role in our life in our life because of these following reasons. The first, in family why we have to... when they don't have enough money to expense, sorry, to pay for expense in our life... in their life. Without money they don't... they don't... they can't buy what they need. For example, they don't... they don't afford all and all other what they need to live comfortable life. Without money their children don't have good conditions in learning, for example good school, good teacher and good book for example. Without money, when they are ill... they... they don't have good doctor, good hospital and can't buy medicine. To do all of these, they need money, have a lot of money. Having money they can buy what they need and all other... they live... they need to live comfortable life. Having money they can have good healthy care... good learning in good condition, and... so they must work hard to earn money. Secondly, in business, if you want to open a shop, found a company,... you build factory, you need money, a lot of money.

Having money you can pay salary for employees or doing business. Third, having money you can do charity... You need money to support poor students... scholarship with.. with the money... with the money you help them... contribute learning... and other. In conclusion, money play... important and necessary role in our life. In our life everybody try to work hard to earn money as much as possible. That's all. Thank you for your listening.

The assessor:Can you buy happiness with money?

Student 9:I think money is... money play role important role in our life, but some... in some extent, money can all what they need, but we can't buy ... spiritual... love and happiness with money.

The assessor:How can you buy it? How can you buy love? You say you can buy happiness with money. OK? But how?

Student 9:I wish I will go abroad to study, in Australia for example. It's my happiness. With money I can do it.

Student 10:

My topic is number 4. If you were born again, would you like to change your sex? If one day a fairy appears... suddenly appears and asks me, 'would you like to change your sex?' I couldn't hesitate to answer, 'no, I wouldn't'... I'd like to be a woman because I find that being a woman... has some advantages... Today I'd like to tell you the reason why I don't want to change my sex. First, being a woman I consider as a fair sex. I can make me more beautiful, more... by making up. You can choose any clothes... fitting my body to make me more beautiful... If I... if I could... all my thought people... people who... around me said that being a woman I'm pleased... If a manusually helps me if I have a problem. Supposed that I'm the last person to go the bus. The bus at that time very crowded and no seat left to be sit on... to be sit on. Yes. Suddenly a man is ready ... to give his seat to me to sit on. If...if...if I'm not a woman, whether the man could... sit her... his to leave or that. Another example, I have just left the supermarket with many things heavy, a man in the street is ready... ready help me by carrying this for me. He want to prove that he is a polite person. But if I were a man at that time the man could carry... I could help him because I'm a girl. To sum up, being a woman I find it very interesting and I don't want to change my sex. That's all. Thank you for your listening.


## Appendix 4: PHIẾU KHẢO SÁT

Thông tin trong phiếu khảo sát này chỉ dùng cho mục đích nghiên cứu. Mong quý thầy cô cho biết một số ý kiến về công tác kiểm tra đánh giá kỹ năng Nói.

1. Khi đánh giá khả năng giao tiếp (taking interactional and transactional short turn) của sinh viên, thầy cô ưu tiên đánh giá những yếu tố nào dưới đây? (đánh số theo mức độ ưu tiên: 1 ưu tiên nhất .....8 ít ưu tiên nhất)

    1. Ngữ pháp đúng.
    2. Phát âm chấp nhận được.
    3. Từ vựng phù hợp.
    4. Giao tiếp được ý tưởng.
    5. Lưu loát.
    6. Làm tốt cả hai vai: người trả lời và người hỏi.
    7. Thường xuyên nói câu dài khi trả lời.
    8. Câu trả lời phù hợp với tình huống liên quan.
    9. Khi đánh giá khả năng trình bày (taking transactional long turns, e.g. oral report) của sinh viên, thầy cô ưu tiên đánh giá những yếu tố nào dưới đây? (đánh số theo mức độ ưu tiên: 1 ưu tiên nhất ..... 9 ít ưu tiên nhất)

    1. Ngữ pháp đúng.
    2. Phát âm chấp nhận được.
    3. Từ vựng phù hợp.

4. Giao tiếp được ý tưởng.

5. Lưu loát.

6. Làm tốt cả hai vai: người trả lời và người hỏi.

7. Thường xuyên nói câu dài khi trả lời.

8. Câu trả lời phù hợp với tình huống liên quan.

9. Nội dung đúng yêu cầu và phù hợp với đề tài.

10. Theo thầy cô để kiểm tra được khả năng nói toàn diện của sinh viên, đề thi nên có số lượng bài tập (task) là…. (có thể có hơn 1 lựa chọn)

1. 1
2. 2
3. 3
4. 4
5. Đánh dấu (  ) vào ô trống cho sẵn loại hình bài tập thi thầy cô cho là phù hợp với năm 1, năm 2 và năm 3.

Năm 1 Năm 2 Năm 3

1. Thảo luận/trao đổi ý kiến giữa 2 sinh viên.
2. Trình bày chủ đề.
3. Phỏng vấn (chỉ trả lời câu hỏi của giám khảo).
4. Đưa ra hướng dẫn để người khác thực hiện công

việc (chỉ đường, vẽ sơ đồ,v.v...).

1. Hoàn thành bài hội thoại bị giấu đi một số lời thoại.
2. Kể chuyện lại sau khi được đọc.
3. Kể chuyện theo tranh.
4. Đóng vai
5. Đọc bài khóa hay một đoạn bài khóa.
6. Những câu hỏi nào dưới đây thầy cô chọn để hỏi sinh viên năm 1, 2 và 3? (Đánh dấu vào ô được chọn)

Năm 1 Năm 2 Năm 3

1. Describe the most important sports event in your

city/country.

1. What do you do to keep fit?
2. What sports do you think are dangerous? Why?
3. What are your suggestions for reducing traffic jams?
4. What can you do to help reduce pollution in the city?
5. What can you think are global problems?
6. What are your suggestions for being a fluent speaker

in a foreign language?

1. What makes a good language learner?
2. How do you learn new words, pronunciation and

grammar?

1. Theo thầy cô phải thực hiện những việc gì dưới đây để tổ chức một kỳ thi Nói?

1. Xác định mục đích của bài thi.
2. Lựa chọn các dạng bài tập/nội dung thi.
3. Tham khảo nội dung chương trình sinh viên đã học.
4. Xác định tiêu chí về lĩnh vực kiến thức, kỹ năng và nội dung cần kiểm tra.

5. Xác định cấu trúc bài thi (gồm mấy bài tập – task).
6. Xác định đặc điểm cụ thể của từng bài tập (task), như mục đích, nội dung kiến thức và kỹ năng, thời gian thực hiện và cách chấm điểm từng bài tập.
7. Theo thầy cô kết quả thầy cô đánh giá khả năng nói của sinh viên có phản ánh đúng năng lực của sinh viên?

1. Chắc chắn đúng
2. Không chắc lắm
3. Không chắc

1. Lý do vì sao thầy cô chọn câu b và c? (có thể có hơn 1 lựa chọn)

1. Sinh viên được chuẩn bị trước các đề tài/câu hỏi thi.
2. Sinh viên dự đoán và chuẩn bị trước các câu hỏi của giám khảo và câu trả lời.
3. Hình thức thi/bài tập (task) không đảm bảo đánh giá khả năng sử dụng tiếng Anh để giao tiếp của sinh viên.
4. Không có tiêu chí, yêu cầu, thang điểm đánh giá cụ thể cho từng loại hình bài tập (task).
5. Xin thầy cô vui lòng cho biết số năm giảng dạy môn tiếng Anh.
6. Thầy cô đã từng tham gia khóa học hay hội thảo nào về kiểm tra đánh giá (testing)?

Rất cám ơn quý thầy cô đã bớt chút thời gian.

QUESTIONNAIRE

Give your opinion on oral language testing.

1. When assessing students' performance on interactional and transactional short turns, which level of priority do you give to the following abilities? (Number your level of priority, e.g. the highest priority is 1 and the lowest 8)

a. Ability to use grammar accurately

b. Ability to pronounce acceptably

c. Ability to use vocabulary appropriately

d. Ability to convey their intended meaning(s)

e. Ability to speak fluently

f. Ability to interact effectively

g. Ability to produce extended speech

h. Ability to make responses appropriate to the situation

2. When assessing students' performance on transactional long turns, which level of priority do you give to the following abilities? (Number your level of priority, e.g. the highest priority is 1 and the lowest 9)

a. Ability to use grammar accurately

b. Ability to pronounce acceptably

c. Ability to use vocabulary appropriately

d. Ability to convey their intended meaning(s)

e. Ability to speak fluently

f. Ability to interact effectively

g. Ability to produce extended speech

h. Ability to make responses appropriate to the situation

i. Ability to make a presentation with an adequate content

3. A test of overall speaking ability should make use of test tasks. (More than one choice is possible.)

a. 1

b. 2

c. 3

d. 4

4. Tick ( ) the elicitation techniques that you think are suitable for students of Year 1, Year 2 and Year 3.

Year 1Year 2Year 3

a. Discussion/conversation

b. Oral report

c. Interview or Question & answer

d. Learner-learner description and re-creation

e. Reading blank dialogue

f. Retelling a story

g. Picture story

h. Role-play

i. Reading aloud

5. Tick ( ) the particular test tasks / questions that you use for students of Year 1, Year 2 and Year 3.

Year 1Year 2Year 3

a. Describe the most important sports event in your

city/country.

b. What do you do to keep fit?

c. What sports do you think are dangerous? Why?

d. What are your suggestions for reducing traffic jams?

e. What can you do to help reduce pollution in the city?

f. What can you think are global problems?

g. What are your suggestions for being a fluent speaker

in a foreign language?

h. What makes a good language learner?

i. How do you learn new words, pronunciation and

grammar?

6. In order to develop an achievement speaking test, which following things must we do? (More than one choice is possible)

a. Identify the test purpose(s)

b. Choose test tasks in the target language use domain.

c. Determine students' topical knowledge and profile of language ability

d. Determine the construct/ability to be measured

e. Determine the structure of the test (the number of test tasks)

f. Identify the specifications of each test task such as purposes of the test task, specified components oral ability to be tested, expected duration of task performance and scoring method

7. Are you sure that the marks you have given on your students' oral test performance can reflect their actual speaking ability?

a. Very sure

b. Not very sure

c. Not sure

8. The reason(s) for your choice of (b) or (c) for Question 7 is/are that. (More than one choice is possible)

a. The students were prepared for the topics or test questions in advance.

b. The students might guess the assessors' questions and prepare in advance the answers to these questions.

c. The test tasks used can't get the students to show their actual ability to communicate in English.

d. There were no criteria and instructions for marking each test task.

9. How long have you taught English? – For ………….. years.

10. Have you ever attended any courses in or workshops on language testing?

Financial maths

# Maths Test Grade 12

Financial Mathematics

Question 1

1. R5000 is invested at 9,6% p.a.interest compounded quarterly. After how many years will the investment be worth R35 000? (5)Question 2Waydene wants to buy a car costing R192 000. She takes out a loan for 5 years with interest charged at 12%p.a. Compounded monthly. 2.1. Calculate the monthly instalments that Waydene will have to pay on the car loan. (5)2.2. After Waydene has paid 45 instalments she decides to settle the balance of the car loan. Calculate the lump sum Waydene will need to pay after she has paid the 45th instament. (7)**Question 3**Determine the time taken in years, for a sum of money to double if the interest rate is 12,6% p.a., compounded half-yearly. (4)Question 4Dudu wants to buy a house for R700 000,00. She has a deposit of R50 000,00 and takes out a loan for the balance at the rate of 18% p.a. Compounded monthly.4.1. How much money must Dudu borrow from the bank? (1)4.2. Calculate the monthly payment if she wishes to settle the loan in 15 years. (4) 4.3. Dudu won a lottery and wishes to settle the loan after the 50th payment. What is the balance. (4)**TOTAL : [30]**

Teaching Science to Young Children:Practical Advice

## Aim

The aim of this module is to offer advice to teachers of children aged 5-7 who want practical ideas to help plan their teaching of science.

## Science is an important subject

A sound knowledge of science is essential in today's modern technical world and will probably be even more so in the future. You should treat science as an important subject. By this I mean you should set aside a set time each week to spend on a science lesson or lessons. You probably already do this for english and mathematics and you should do the same for science. How much time you spend each week is up to you or your education board but I would recommend that you spend one hour a week in a formal science lesson and perhaps another hour a week as part of topic work.

## Science is a Practical Subject

Science, in essence, is a way of finding things out about the world. The scientific method, has allowed man to make very rapid advances in knowledge and understanding about the natural world. This method is best demonstrated by Gallelao's experiments with gravity. The wizdom at the time suggested that heavier objects fell faster than light ones. He dropped two masses and timed how fast they fell to disprove the accepted wizdom. Such a simple experiment but no one had thought to do it before him.

When teaching science aim not to teach facts, instead aim to teach skills. Pupils with the necessary skills will, with your help of course, be able to find out the facts for themselves. The skills needed to become a scientist can only be learned through practice. Pupils gain theses skills by performing experiments. You should aim to do at least one investigation or experiment every week. For very young children, the experiments should be simple with clear results.

A very simple experiment for a child in this age group might be to find out the answer to "What happens to an ice cube if we leave it on a dish in the classroom?". This is an experiment that a child as young as three or four can comprehend. As the child grows older related questions can be investigated. How long does it take for the ice cube to completely melt? Does the position it is placed in the room affect how fast it melts and if it does then what can that tell us about the temperature around the room.

## Science is a playful subject

Children, especially young children, are naturally curious about the world and will, if left to their own devices find out about the world through play. As teachers we can exploit this natural tendency and encourage playfulness as a powerful aid to learning. Allow pupils time to play with science equipment. You could for example, give them a torch and ask them to find out as many things as they can about shadows, they must report back to the rest of the class in 5 minutes. If you do this you will find them teaching themselves. You can sit back and simply enjoy the fun.

Formal games can also work well. You want them to learn the functions of different parts of a plant? Make a set of cards with "makes food for the plant", "soaks up water from the soil" and so on. Give each child a picture of a plant to colour in and explain the rules. Each child takes a card from the pack. They identify the plant part from the function then colour it in. If they get a card they had before too bad. If they colour in the wrong part they are out. First child with a fully coloured in plant is the winner. It does take some time and planning on your part because you have to make up the cards but once you do you can reuse them year after year. Again you will be able to sit back and watch the children teach themselves and each other! Other games, such as bingo, dominoes, and snakes and ladders type games can all be adapted to help teach science.

## Science is a fun subject

Pupils love doing hands on work. With a little effort you can make science their favourite subject. Plan your lessons carefully. Keep learning objectives clear and simple. Do not try to teach too much all at once. Plan multiple

activities that have the same learning objectives. Children have a short attention span, keep them enthusiastic by not exceeding that span. An hour long lesson should be divided up. For example you might plan :

- A 10 minute starter activity (such as a video) to get them interested.
- Two 15 minute practical activities to get them learning.
- A 15 minute game to reinforce the learning.
- A five minute teacher led whole class discussion to recall what they have just learned.

The time will fly by, fun will be had by all, and the pupils will come to love science.